THE MEASUREMENT OF SILENT READING

SCALE FOR MEASURING ABILITY IN SILENT READING

1. This naughty dog likes to steal bones. When he steals one he hides it where no other dog can find it. He has just stolen two bones, and you must take your pencil and make two short, straight lines, to show where they are lying on the ground near the dog. Draw them as quickly as you can, and then go on.

5. Have you ever seen such a strange bird? He is hard to find because he sleeps in the woods during the day and does not come out until night. Take a pencil and tell people what the bird's name is by writing the word OWL, with a capital O, under the books on which the bird is standing.

9. Here is a Christmas pudding with five lighted candles at each side. You must take your pencil and make five little lines to stand for the five matches that were used to light these ten candles. Put three of them on one side by one candlestick, and two of them on the other side by the other candlestick.

13. This butterfly has been living in a warm cocoon, but now he has come out and is flying around exploring the world. He is beginning to grow tired, so you must draw a little stick under his feet on which he is resting, and another one above him to which he can easily fly.

17. This happy man works in an office and is carrying a great many papers to the boss's desk. He has just dropped two of them, and you must draw them lying on the floor. To do this make one small square on the floor in front of him and another one lying on the floor behind him.

2. This man is an Eskimo who lives in the far north where it is cold. There has just been a big storm, and all the ground is white with snow. The man has been walking and has made many footprints in it. With your pencil quickly make four of these in the snow just behind him.

6. This small chap is afraid to start for school. The teacher will scold unless he brings his books, but the big owl is sitting on them. Grasp your pencil bravely and cross the owl out of the previous picture with two black lines, so that the child can rescue his belongings. Remember not to use more than just two lines.

10. Help this gay and lively young lady to have a happy time playing all by herself by taking your pencil and drawing a skipping rope with the two ends held in her two hands, and make the rope so that everyone can see that it is just passing beneath her feet as she skips over it.

14. This man is playing in a bowling alley at his club. Each player rolls two balls, one after the other. This man has already rolled one and you may make a little circle in front of him to show where it went, and after that make two more just behind him for the next player to use.

18. This man is blowing soap bubbles with his long pipe, and he has three bowls of soapy water on the rug beside him. Draw one round bubble still fastened to the upper end of his pipe, and after you have done so, draw two more floating in the air in front of him.

3. This book is lying on the desk, but it is hard to make it stay open. With your pencil draw a single straight line to represent a ruler lying across the book to hold the pages open. Be sure to make the line from one side to the other, across the book, instead of making it go up and down.

7. These two flags are used as signals to give notice of changes in the weather. The white flag means fair, so you may now take your pencil and make a capital F under the white flag, to stand for fair. The blue flag means storm, so make a capital S under the blue one.

11. Here is another picture of the little girl who owns the skipping rope. This time she has a big hoop in one hand, but you may make a picture of her skipping rope with one end held in her hand and the other end dragging on the ground behind her, since she cannot use it while she has her hoop.

15. This proud old eagle with spreading wings is in a very risky place on a big smooth glass ball. Please keep the ball from rolling and upsetting the eagle by taking your pencil and quickly drawing one large stone close against the ball on one side and two little stones close against it on the other side.

19. When the road is rough the porter finds it hard to push this wheel chair. Draw a line to show where the road is. Be sure to make the line in front of the chair smooth so that the chair will roll along easily, but make the line in back of it uneven because up to this time the path has been rough.

4. This savage Indian is going to war, as you can tell because he wears a war bonnet trimmed with eagle feathers. Three of the feathers have fallen out, and you must quickly draw them lying on the ground behind him. One of them is very near him and the other two are lying side by side farther off.

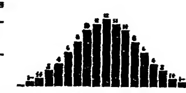
8. A man is riding in this covered chair. He does not want to be seen; and you may take your pencil and blacken the windows so that no curious person can peek in. Then blacken the lower part of the chair and the handles so that it will look as if the whole chair was painted black. Work quickly.

12. The children will soon find the gifts Santa Claus left in their stockings. Since the littlest girl will wonder what is in hers, draw a hole as fast as you can in the foot of the littlest stocking so that she can peek in, but be careful not to make any in the toes of the three other stockings.

16. This new weather flag shows that cold weather is coming. To make all the flags on this page tell the truth go back to the white flag that you marked with an F and make a little black square in the middle of it. When you have done this cross out the letter F below it that you made earlier.

20. Years ago children learned in school to make fancy letters that were pretty but hard to do. This one is an M and you may show how much more sensible our present methods are by making a printed capital M on one side of this picture and a written capital M on the other side of it.

CREDIT CORRESPONDING TO EACH NUMBER OF PARAGRAPHS MARKED IN EACH GRADE



THE MEASUREMENT OF SILENT READING

MAY AYRES BURGESS

DEPARTMENT OF EDUCATION
RUSSELL SAGE FOUNDATION
130 EAST TWENTY-SECOND STREET
NEW YORK CITY

Printed February, 1921, 2000 Copies
Reprinted May, 1921, 2000 Copies

WM. F. FELL CO. PRINTERS
PHILADELPHIA

TABLE OF CONTENTS

LIST OF TABLES AND DIAGRAMS

THE MEASUREMENT OF SILENT READING

CHAPTER I

THE MEASUREMENT OF READING

When the United States declared war against Germany, four million picked men were chosen for military and naval service. Many thousands of those in the Army were tested by the Psychological Section of the Surgeon General's office. Two types of tests were given: the "Alpha test," which required the ability to read and write, and the "Beta test" for foreigners and illiterates. Those who failed on Beta were subjected to individual testing. Since about one-quarter of the men were judged unable to take the Alpha test, the results indicate that, if those examined were fairly representative of all, there must have been over one million of our soldiers and sailors who were not able to write a simple letter or read a newspaper with ease.

Lack of Schooling Not Wholly Responsible

These findings have attracted widespread attention and comment; but there is another significant fact concerning them that has not been emphasized.

This is that although one-fourth of the men could not read well enough to take tests based on reading, this deficiency was not caused by their never having learned to read. The fact is that an overwhelming majority of these soldiers had entered school, attended the primary grades where reading is taught, and had been taught to read. Yet, when as adults they were examined, they were unable to read readily such simple material as that of a daily newspaper.[1]

Probably no more striking evidence could be secured of the serious and little realized fact that many people, in spite of going to school, never really learn to read easily. They do not acquire sufficient ability in reading to use it freely as a tool. What happens to a child who learns to read, but no longer retains the facility when he becomes an adult, is something like what happens to many high school and college graduates in the matter of French and German. They have learned to pronounce and translate from the pages of their French and German textbooks, but after leaving school they never buy a French story or read a German periodical.

It may also be noted that some years before the outbreak of the European War comments appeared in the educational magazines upon the findings of the

[1] These data are secured from the records of the Surgeon General's office, showing results of psychological examinations, and the school records of 12,000 hospital patients. See also Baldwin, Bird T., Distribution of school training of wounded soldiers, in *School and Society*, Vol. X, No. 258, Dec. 6, 1919, p. 680.

French military authorities who tested the reading ability of recruits entering the Army in one year under the compulsory military service system. The results secured were closely similar to those found in the recent work among American soldiers. It was reported that many thousands of the young Frenchmen, who had attended the elementary schools during the compulsory attendance period, were not able, upon reaching the age of 20, to use their reading ability as an ordinary, everyday tool. Like our own soldiers, they would be counted by the census as literates; but their literacy was of such low grade as to be of little help for ordinary reading purposes.

In the United States, reading has long been recognized as the first important subject for a child to learn. It is in fact the most important school subject that he ever will learn; for the ability to read opens the doors to all other fields of human knowledge. It makes possible communication with others at a distance. It makes available the results of other people's experience, and the conclusions of their thinking. The modern interest in education is casting out old subjects from the curriculum, and introducing new subjects to take their places; but with all the changes, made and contemplated, reading maintains its supremacy as the most important single subject the child can learn.

The fact that many children never really learn to read is not due to indifference on the part of the teachers. It is due rather to the fact that, although a large proportion of the time of the elementary

grades is devoted to reading, educators have not been able either to measure the results secured, or to make diagnoses in cases of individual difficulty. They have been unable to tell, on the one hand, which methods of teaching were successful and which were not; or, on the other, what influences were operative in preventing children from becoming good readers, and how these influences might be overcome.

Beginnings of the Modern Movement

The inception of the modern movement for scientific measurement in education dates from 1910 and was marked by the publication of the first of the modern scales for the measurement of classroom products.[1]

This earliest scale was a device for measuring the qualities of samples of handwriting in numerical terms. The scale itself was a sheet of paper on which there were reproduced samples of children's handwriting, ranging from those that were of such poor quality as to be illegible up to other samples of progressively better quality, until finally, at the upper end of the scale, there were found reproductions of handwriting of substantially perfect, or copperplate, quality.

In the monograph which accompanied his scale, Dr. Thorndike remarked that, previous to that time, educators had been in the same condition with respect to handwriting as were students of tempera-

[1] Thorndike, E. L. Handwriting. *Teachers College Record*, 2: March, 1910, 1–93.

ture before the discovery of the thermometer. In that early day it had not been possible to measure ordinary temperatures beyond the cold, cool, warm, hot, and very hot, of subjective opinion. Similarly it had, before 1910, been impossible to measure the quality of handwriting except by such vague standards as that one's personal opinion was that a given sample was very bad, bad, good, or very good, etc.

This earliest scale was rapidly followed by others for measuring the classroom products in different subjects and by numerous reports of extensive applications of these new educational adjuncts. The movement has spread so rapidly that now, at the end of its first decade, there are in existence more than a hundred standardized tests and measuring scales, and over a thousand reports on the results secured by using them.

Need For Measurement In Reading

The object of these measurements is to make it possible to study education by finding out what the children can do. These new methods make the child and not the teacher the center of interest. They proceed by measuring the accomplishment of the pupil, rather than by analyzing the methods of the teacher.

Measurements of this sort, that can be easily administered and readily interpreted, are peculiarly needed in reading. The recent army tests have furnished impressive evidence, on a large scale, that results of school work in reading need to be improved. Such improvement would be greatly facilitated by

better methods for judging results of classroom work; and this fact is clearly indicated by the advances that have already resulted from the use of scales and tests for handwriting, arithmetic, and spelling. Much work has been done in the measurement of reading, but the inherent complexities of the task have resulted in tests that are, for the most part, harder to administer and far more difficult to interpret than those generally used for writing, spelling, and arithmetic.

The Plan of this Book

During the past year, the Department of Education of the Russell Sage Foundation has attempted to devise a new scale for the measurement of silent reading. The aim of the present volume is to describe the new scale, Picture Supplement Scale 1, as it was finally developed; to relate the experiments upon which it is based; and to give a brief account of the principles which seemed to be involved in its construction. The plan of the book is as follows:

Chapters 2 and 3 describe the new scale which has been adopted and the five other scales which preceded it and were discarded.

Chapters 4 to 9 deal with the principles of measurement which came to be recognized as fundamental in any attempt to measure reading.

Chapters 10, 11, and 12 recount the statistical procedures followed in making the test, assigning scale values, and judging the reliability of the results of the new scale, Picture Supplement 1.

Summary

1. Tests of American soldiers during the war showed that about one-fourth of them were unable to read newspapers easily or write simple letters. Investigations made some years earlier by the French military authorities showed that similar conditions existed among recruits entering the French army.

2. Most of these soldiers in both armies had attended school during their boyhood and learned to read. They had not retained the ability well enough to use reading as an everyday practical tool in adult life a few years after leaving school. This shortcoming in our educational methods and results needs remedy.

3. Reading is the most important single subject the child has to learn. Poor results of schooling are attributable not to lack of attention but to methods which fall short of being fully effective.

4. The teaching of reading could be rendered more effective if good measuring instruments were available to show which teaching methods produce the best results; what influences are operative in preventing children from becoming good readers; and how those influences may be overcome.

5. The plan of the present volume is as follows: Chapters 2 and 3 describe the new scale, and the five different scales which preceded it. Chapters 4 to 9 deal with the principles of measurement involved. Chapters 10, 11, and 12 recount the statistical procedures followed in making the test and scale of PS-1.

CHAPTER II

PS-1, A PICTURE SUPPLEMENT SCALE

In measuring the child's handwriting, drawing, or composition, the teacher asks, "How good is it?" In measuring arithmetic she asks, "How many did he get right?"; but when reading is under consideration another and different inquiry arises, and the teacher asks, "How much did the child get out of what he read?"

How Much Did He Get Out Of What He Read?

Every scale for measuring silent reading is an attempt to answer this question by means of having the child do something which he can do correctly only if he understands the material given him to read. The commonest method employed has been to have the child read a selection to himself and then reproduce it orally or in writing. Thus Mr. Brown[1] and Mr. Courtis[2] have the children read a long connected story and then write as much of it as they can remember. Mr. Don C. Bliss of Montclair, New Jersey, in an un-

[1] Brown, H. A. The Measurement of Ability to Read. *New Hampshire Dept. of Public Instruction Bureau of Research, Bulletin 1*, p. 57.

[2] Courtis, S. A. *Courtis Standard Tests, Silent Reading, Test No. 2*, Detroit, Michigan.

published study, Mr. Kallom,[1] Dr. William Gray,[2] and Dr. Starch,[3] present several short stories or selections and have them reproduced in the same way. The child's score depends upon the per cent of all the ideas which he can correctly remember and reproduce. Mr. Bliss and Mr. Courtis make an additional measurement by counting not only the ideas reproduced but the actual words remembered and used by the child in his written account.

Other students have held that measurements such as these confuse ability in reading with ability in English composition. To overcome this difficulty Dr. Gray and Mr. Kallom have added to their tests supplementary questions to be answered by the children. This same method is also followed by Mr. Fordyce[4] and Mr. Adams.[5] They answer the question "What did the child get out of what he read?" by saying "As much as he can reproduce in response to questions about it."[6]

[1] Kallom, Arthur A., Boston, Dept. of Educational Investigation and Measurement. *Standards in Silent Reading, Bull. No. 12, School Document 18,* 1916.

[2] Gray, W. S. *Gray's Reading Tests: Silent and Oral.* University of Chicago, Chicago, Illinois.

[3] Starch, D. *Starch's Silent Reading Tests.* University of Wisconsin, Madison, Wisconsin.

[4] Fordyce, Charles, Teachers College, University of Nebraska. *A Scale for Measuring the Achievements in Reading.* University Publishing Co., Chicago and Lincoln, 1917.

[5] Adams, William C., State Normal School, Plymouth, N. H. *Silent Reading Tests.* Ed. E. Rabb & Co., Boston, 1920.

[6] For an exceptionally full and helpful discussion of the problems connected with measurements of oral and silent reading, see the monograph by Dr. William S. Gray, already referred to in the second footnote.

In a series of experiments recently conducted by the School of Education of the University of Chicago, evidence has been produced which indicates that the question method of testing is superior to that resting upon unassisted reproduction because, by utilizing a fundamental law of memory and recall, it tends to disassociate failures of memory from failures in reading.

A child who has read and understood a story is often unable to write a connected account of it, because his ability to recall ideas in series is limited to a span much more brief than that required by the story. A score which depends upon the number of ideas reproduced will, in such a case, be very low. It cannot properly be taken as a measure of his reading ability, because it largely depends upon a very different ability, that of being able to reproduce ideas in sequence.

The good reader understands and notes each idea as it is presented; and the location of the idea in the series seems to make little difference in the vividness with which it is impressed upon him. He is as well able to remember ideas near the end of a story as those near the beginning, or in the middle. If the story is so constructed that such a procedure seems to him logical, the child can start at the end of a story and work backwards just as far and as efficiently as he can when he starts at the beginning and works forward. Moreover, he can start at the middle and work in either direction.

What he cannot do, is to go very far in the process

of recall from the point where he has started. The ability to remember points in a series runs in "takes," or short assignments. The child starts readily, recalls several points in their proper sequence, and begins to falter. If no help is given, he stops; but if he is asked a question he is again stimulated. The question acts as a key which opens the way to another consecutive series of ideas. He remembers these; and, the impulse being over, again falters. Where questions are so arranged that they accord with the normal recall-spans for the children being tested, they greatly relieve the burden upon memory and assist in removing the alien memory element from scores intended to measure ability in reading.

Most of the scales and tests for measuring silent reading have been based upon the principle of required reproduction, with or without questions; but the Thorndike Alpha,[1] Kansas Silent Reading test (Kelly),[2] Kansas Standardized Silent Reading test,[3] part of the Courtis test, and the Haggerty-Noonan test[4] seek to eliminate the demands upon memory and ability in composition by allowing the children to reread the material upon which they are being tested,

[1] Thorndike, E. L. *Thorndike's Scale Alpha for Measuring the Understanding of Sentences.* Teachers College, Columbia University.

[2] Kelly, F. J. *The Kansas Silent Reading Test.* Kansas State Normal School, Bureau of Educational Measurement, 1915.

[3] Monroe, W. S. *Standardized Tests in Silent Reading,* Kansas State Normal School, Bureau of Educational Measurement.

[4] Haggerty, M. E., and Noonan, M. E. *Achievement Examination in Reading,* World Book Company, 1920.

and by calling for responses based on the text and involving a minimum of writing. They seek to answer the question "How much did the child get out of what he read?" by finding out how well he is able to answer questions or obey orders based on the reading material. In general it may be said that existing scales and tests of silent reading seek to answer the teacher's question in three ways: The child got out of his reading "enough to reproduce," or "enough to answer questions," or "enough to follow directions."

Limitations of Existing Scales

Reference has already been made to the limitations of certain of the existing scales, and the efforts that have been made to overcome them. Most of the tests and scales for measuring silent reading suffer from one or more of four important limitations, which have been pointed out, not only by teachers who have used the tests and scales in their classrooms, but by the authors of the measuring instruments themselves. While comments are made in many forms, they may be briefly summarized under four groups, as follows:

In the first place, the statement is frequently made that the tests and scales in question, while endeavoring to measure ability in reading, actually measure in addition other and different abilities. Those which utilize the scheme of having the children read a story and reproduce it are held by some to measure not primarily the ability to read, but rather the ability

to remember, to write English composition, to discriminate between words and phrases actually used and their equivalents, or to answer questions. The tests and scales which present a number of brief paragraphs and have the children answer questions about them or follow instructions given by them, are spoken of as measuring the ability to reason correctly, to infer, to remember, to do arithmetic, to solve puzzles, to resist irrelevant suggestions, to detect absurdities, to make ethical judgments, to discriminate between the meanings of words, to think in terms of spacial relations, to cross out certain letters in a series, and so on, through an infinite number of abilities, all of which are often found in conjunction with the reading process, but none of which can properly be called the ability to read.

A special case of this first type is frequently found in tests which are constructed in steps of increasing difficulty, so that the child starts at the beginning and works up through the series as far as he can go within a definite time. In tests of this sort, the final score which is assumed to measure the hardest work which the child can do, actually measures instead, the hardest work he succeeded in doing in the time he was allowed to try. It is held, that is, that scales for difficulty reached in a given time, result in mongrel scores which are combinations of the difficulty the child reached, the speed with which he worked, and the accuracy of the work he did. The attempts to equate time and accomplishment are frequently held

to be invalid and unreliable for purposes of comparative measurement.

The second fundamental limitation of existing measuring instruments, particularly of those tests and scales which consist of a series of different tasks, either of ascending difficulty or of equal difficulty, is that the different tasks are not consistent, in that they do not conform to any single set of conditions for testing. In some instances they measure a series of different abilities; and the final score is a conglomerate of them all. In other cases, while the ability to be measured is approximately the same throughout all the tasks, the conditions under which the child works are markedly different. One section calls for a single response and another for several responses. One requires a few seconds to solve correctly, while another involves several minutes of rapid thinking and working. Again, sections vary in length from a single sentence to half a page; some are in small type, some in large; some have pictures, some do not. Inconsistency in the component sections of the testing material makes for unreliable scores which cannot readily be analyzed or interpreted.

The third objection commonly found against tests and scales for measuring silent reading is that they are difficult to administer and score. Some of them require a full classroom period or more. Others are to be given for so short a time that a stop-watch is necessary if the conditions of the test are to be strictly complied with. Still others require carefully timed individual testing, so that their use for a

large class is practically out of the question. Again, most of the tests are long and difficult to score. They require keys to which the teacher can refer, to find what the correct answer should be; or they demand the counting of words, judgment as to the proportion of ideas reproduced, and so on. In most cases, scoring the results is so long and hard a task that it proves a heavy burden upon the busy classroom teacher.

Finally, most of the existing tests and scales cannot be used for comparing the achievements of individual children with the achievements commonly found for other children of like amounts of maturity and training. Scores as gathered are frequently thrown together, so that it is impossible to separate the records of third grade children from those of the fourth grade, fifth grade, and so on. Again, scores are presented in terms of the numbers of tasks correctly done, but no information is given as to how these numbers rank in comparison with the numbers usually found for each given grade. Scores are not turned into their equivalent scale values for the separate grades, and it is therefore impossible to compare the relative standing of one child with the standing of most other children who belong to the same grade and have therefore presumably been given the same opportunity to learn as has he.

These, then, are the four limitations commonly recognized for the existing tests and scales for measuring silent reading. They are, first, that the instruments in question measure not only reading

ability, but other abilities widely different from it; second, where such tests and scales consist of separate tasks in a series, these tasks are not consistent in character; third, most of the tests and scales are difficult to administer and hard to score; and fourth, information is frequently lacking whereby the achievements of an individual child can be compared with the achievements of other children in his own grade.

Reading For Practical Purposes

The new silent reading scale, Picture Supplement Scale 1, is an attempt to devise an instrument which shall be free from these four fundamental limitations. It is designed to measure silent reading ability by strictly utilitarian standards. The general scheme is to present a series of pictures and paragraphs about them. These paragraphs consist of instructions which the pupil follows by marking with his pencil a line or lines to supplement the picture. His ability to do this in accordance with the printed instructions reflects the rapidity and accuracy with which he can read.

The aim of such a test is to find out how much printed material of a given level of difficulty the child can read "well enough for all practical purposes." The attempt is to devise a test in which the child can readily succeed if he reads well enough to grasp the important thought in each section, and in which he cannot succeed at all unless he does comprehend each important thought. This is the interpretation which has here been put upon the phrases "utilitarian read-

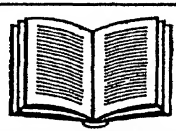
ing," and reading "good enough for all practical purposes." The attempt has been made to keep the test of a uniform level of vocabulary and phraseology, and of a uniform level of thought difficulty, and then find out how much of the kind of reading involved the child can do within a given amount of time; which in this case has been fixed at five minutes.

The scale itself is a single sheet of paper 12 inches wide and 19 inches long. The sheet is divided into five columns. Each column is divided into four sections, and in each of these sections there are a picture and a paragraph about it. The instructions are extended though the paragraphs in such a way that they cannot be fully grasped unless the entire paragraph is read. They are so worded that they cannot be misunderstood in moderately careful reading; and can be correctly followed in only one way. The child who guesses is almost sure to make a mistake; but if he reads carefully his answering markings will be "right"; that is, they will be in accord with the instructions given him. His score is the number of paragraphs which are marked "right."

This picture supplement scale has been given the office designation of PS-1. The five columns of the scale, reduced to two-thirds of their actual size, are reproduced on the five following pages.

1. This naughty dog likes to steal bones. When he steals one he hides it where no other dog can find it. He has just stolen two bones, and you must take your pencil and make two short, straight lines, to show where they are lying on the ground near the dog. Draw them as quickly as you can, and then go on.

2. This man is an Eskimo who lives in the far north where it is cold. There has just been a big storm, and all the ground is white with snow. The man has been walking and has made many footprints in it. With your pencil quickly make four of these in the snow just behind him.

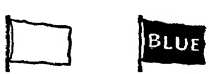


3 This book is lying on the desk, but it is hard to make it stay open. With your pencil draw a single straight line to represent a ruler lying across the book to hold the pages open. Be sure to make the line from one side to the other, across the book, instead of making it go up and down.

4. This savage Indian is going to war, as you can tell because he wears a war bonnet trimmed with eagle feathers. Three of the feathers have fallen out, and you must quickly draw them lying on the ground behind him. One of them is very near him and the other two are lying side by side farther off.

5. Have you ever seen such a strange bird? He is hard to find because he sleeps in the woods during the day and does not come out until night. Take a pencil and tell people what the bird's name is by writing the word OWL, with a capital O, under the books on which the bird is standing.

6. This small chap is afraid to start for school. The teacher will scold unless he brings his books; but the big owl is sitting on them. Grasp your pencil bravely and cross the owl out of the previous picture with two black lines, so that the child can rescue his belongings. Remember not to use more than just two lines.



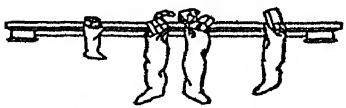
7. These two flags are used as signals to give notice of changes in the weather. The white flag means fair; so you may now take your pencil and make a capital F under the white flag, to stand for fair. The blue flag means storm; so make a capital S under the blue one.

8. A man is riding in this covered chair. He does not want to be seen; and you may take your pencil and blacken the windows so that no curious person can peek in. Then, blacken the lower part of the chair and the handles; so that it will look as if the whole chair was painted black. Work quickly.

9. Here is a Christmas pudding with five lighted candles at each side. You must take your pencil and make five little lines to stand for the five matches that were used to light these ten candles. Put three of them on one side by one candlestick, and two of them on the other side by the other candlestick.

10. Help this gay and lively young lady to have a happy time playing all by herself by taking your pencil and drawing a skipping rope with the two ends held in her two hands, and make the rope so that everyone can see that it is just passing beneath her feet as she skips over it.

11. Here is another picture of the little girl who owns the skipping rope. This time she has a big hoop in one hand, but you may make a picture of her skipping rope with one end held in her hand and the other end dragging on the ground behind her, since she cannot use it while she has her hoop.



12. The children will soon find the gifts Santa Claus left in their stockings. Since the littlest girl will wonder what is in hers, draw a hole as fast as you can in the foot of the littlest stocking so that she can peek in, but be careful not to make any in the toes of the three other stockings.

13. This butterfly has been living in a warm cocoon, but now he has come out and is flying around exploring the world. He is beginning to grow tired; so you must draw a little stick under his feet on which he is resting, and another one above him to which he can easily fly.

14. This man is playing in a bowling alley at his club. Each player rolls two balls, one after the other. This man has already rolled one and you may make a little circle in front of him to show where it went, and after that make two more just behind him for the next player to use.

15. This proud old eagle with spreading wings is in a very risky place on a big smooth glass ball. Please keep the ball from rolling and upsetting the eagle by taking your pencil and quickly drawing one large stone close against the ball on one side and two little stones close against it on the other side.

16. This new weather flag shows that cold weather is coming. To make all the flags on this page tell the truth go back to the white flag that you marked with an F and make a little black square in the middle of it. When you have done this cross out the letter F below it that you made earlier.

17. This happy man works in an office and is carrying a great many papers to the boss's desk. He has just dropped two of them, and you must draw them lying on the floor. To do this make one small square on the floor in front of him and another one lying on the floor behind him.

18. This man is blowing soap bubbles with his long pipe, and he has three bowls of soapy water on the rug beside him. Draw one round bubble still fastened to the upper end of his pipe, and after you have done so, draw two more floating in the air in front of him.

19. When the road is rough the porter finds it hard to push this wheel chair. Draw a line to show where the road is. Be sure to make the line in front of the chair smooth so that the chair will roll along easily, but make the line in back of it uneven because up to this time the path has been rough.

20. Years ago children learned in school to make fancy letters that were pretty but hard to do. This one is an M and you may show how much more sensible our present methods are by making a printed capital M on one side of this picture and a written capital M on the other side of it.

At the bottom of the sheet, underneath the testing material, are instructions to the teacher for giving the tests, marking the papers, and giving credit for each number of paragraphs right. A table is furnished by which scores can be turned into credit marks for each grade from the third through the eighth, and the distributions of credits usually found in those grades are shown by text and diagram. The instructions for giving the test, marking the papers, and assigning credits, and the statement of results usually found, are reproduced below. Following them are reproductions of the scale-table, and the diagram showing the typical grade distribution of children according to their silent reading ability.

GIVING THE TESTS.—1. See that each child has a pencil and that the teacher has a watch. 2 Distribute scales, face down. 3. Have children write on backs of sheets their names, grade, and date. 4. Tell children they are to have a test in reading. Hold scale up and explain that each paragraph tells them to do something to the picture above it with their pencils. They must read carefully, to make sure just what they are to do. They are to read and mark the paragraphs in order, starting at the top and working down, through the first, second, third, and so on. They must do as many as they can in five minutes. 5. Make sure that the pupils understand, then tell them to turn papers over and begin. Allow exactly five minutes. Collect papers.

MARKING THE PAPERS.—Count every paragraph correct in which the marking, no matter how crude it may be, exactly follows instructions. Count every paragraph wrong in which the marking does not exactly follow instructions. Remember this is a test of reading, not of drawing. The pupil's score is

the number of paragraphs correctly marked. Write this number at the top of the paper.

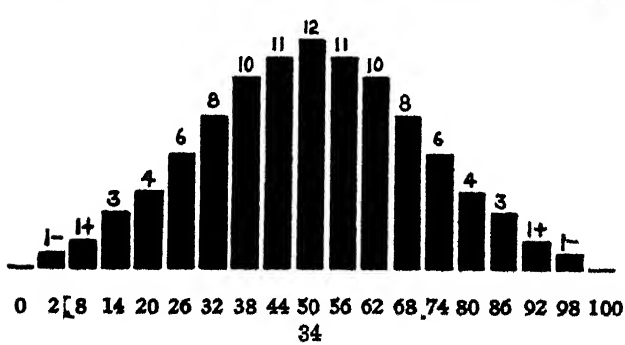
GIVING CREDIT.—The credit to be assigned in each grade for each number of paragraphs correctly marked is shown in the table at the right. A third grade child having ten paragraphs right should be marked 80, a fourth grade child having seven right, 50, and so on Credits in table are for February 1. To adjust for other periods add or subtract from each child's mark as follows:

Grade 3	Oct. 1, +6	Dec. 1, +3	April 1, −3	June 1, −6
Grades 4–8	+2	+1	−1	−2

Give no marks less than 0 or more than 100. Write the credit the child receives at the top of his paper.

CREDIT CORRESPONDING TO EACH NUMBER OF PARAGRAPHS MARKED IN EACH GRADE

Grade	Number of paragraphs read and marked correctly																			Grade
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
3	0	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	..	..	..	..	..	3
4	0	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	..	..	..	4
5	0	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	..	..	5
6	0	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	..	6
7	..	0	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	7
8	..	..	0	2	8	14	20	26	32	38	44	50	56	62	68	74	80	88	92	98	100	8



RESULTS USUALLY FOUND.—In the diagram at the right of the table, columns show per cent of children in the average grade receiving each mark from 0 to 100. Figures below columns show marks or credits, and figures above columns show per cents of children commonly receiving those credits Thus 12 per cent usually receive a mark of 50, eight per cent one of 68, and so on. The lowest third in the average class receive marks of from 0 to 38; the middle third, from 44 to 56, and the best third, from 62 to 100.

A SCALE FOR AMOUNT DONE

The new scale for measuring silent reading, Picture Supplement Scale 1, has four outstanding characteristics. The first of these is that it makes a definite attempt to measure a single ability, and that is the ability to read silently a single type of material, at a constant level of difficulty, in a fixed period of time. It measures the amount of reading of a practically useful nature which the child can do in five minutes.

The amount of such reading is the important thing to measure, because, when grades of performance are equal, the difference in speed amounts to the difference in efficiency. This fact is illustrated by common experience. Of several typists who are equally accurate in copying, the one who turns out the most pages in an hour is the best worker. Again, on a newspaper, where two reporters write equally accurate and interesting stories, the one who gets his copy to the editorial desk with the greater speed is the one who receives the harder assignment. Of two tennis players who can place the ball with equal accuracy on the other side of the net, the one with the more speed wins.

In these, as in most other walks of life, efficiency means "the ability to get it right the first time."

In reading there is almost always a demand for speed. In the early discussions of rates of silent reading, high school and college teachers reported cases of students who, although of excellent standing in most of their subjects, were making poor records in classes where large amounts of supplementary reading were assigned. It was found that, with these students, the rate of silent reading was often so slow that they were genuinely unable to keep up with the pace easily set by their classmates. There are students who read the ordinary novel at the rate of 15 pages an hour; and there are others, often in the same classroom, who can read the same novel at the rate of 150 pages an hour. It is, moreover, true that while there are exceptions to the rule, rapid readers are usually those who get most out of what they read; in general, speed and comprehension increase together. In emphasizing speed, then, we are indirectly emphasizing comprehension; and the measure which is of greatest importance to those who are teaching reading is that which answers the question, "How much can he get, right, how fast?"[1]

Controlling Factors Kept Constant

The second characteristic of the new scale is that a distinct attempt has been made to locate and recog-

[1] For a discussion of rates of careful and normal reading, see Courtis, S. A., Standards in Rates of Reading, 14th Yearbook, National Society for the Study of Education, Part 1, 1915, pp. 44–58.

nize the factors which control the results secured in silent reading; and, with the exception of the speed with which the child reads, to maintain them constant throughout the testing, so that their influence will not enter into the differences of the results secured.

The controlling factors in silent reading are numerous. The following list shows 25 which were especially studied and kept in mind in making the present tests; and upon inspection it will be seen that each of these 25 might be divided into sub-sections, extensions, and ramifications, in endless variety.

Controlling Factors in Silent Reading

To be measured

Amount child can do in given time

To be eliminated

Complex thought
Abstract thought
Technical thought and language
Catches
Puzzles
Accidental leads
Demands for spacial imagination
Irrelevant dramatic appeal
Ability to reproduce
Ability to remember
Ability to reason, or infer
Involved style

To be held constant throughout test

Memory span requirements
Attention span, multiple strains
Difficulty of action demanded

Time required for complying with instructions
Vocabulary difficulty
Sentence structure
Word arrangement
Amount of material to be read
Uniformity of print
Uniformity of space relations between pictures and print
Ease of finding place on paper
Interest and corresponding effort on part of child

Of this long list of controlling factors, one, the time required by the child to read a paragraph correctly, was adopted as the variable to be measured; and the other 24 with their various subdivisions were, in so far as was practically possible, retained at constant levels throughout each paragraph of the test.

Thought, Vocabulary, and Style

In general, an examination of the scale as finally produced will show the method adopted for treating each of the 25 controlling factors which were especially considered in the construction of Picture Supplement Scale 1. It is worth while, however, to make special comment on a few of the more important points, in order to indicate the sorts of standards which were selected to guide the construction of the testing material.

For example, every paragraph consists of directions for doing something with a pencil to the picture above it, and the things which the child is asked to do are always simple, and always equally simple. Remarks which do not bear on the task required have

been ruled out, so that, as far as possible, no conflicting interests enter. The thought is always of the same kind, and always simple. Instructions which call for technical thinking are eliminated. There are no puzzles or catches, and the paragraphs have been carefully revised after testing, to remove accidental ambiguities. Care has been taken to emphasize the particular points in the instructions on which successful scores depend.

Pictures which are uninteresting have been discarded, and so also have the pictures or paragraphs so especially interesting that they are apt to start the child's imagination working and make him forget what he is required to do. The attempt has been made to write all the paragraphs in such a way that the ordinary child will enjoy following the directions of each one and will be eager to attempt the next.

The vocabulary used is that of ordinary newspaper English. Long, strange, and technical words are ruled out; and most of the words used are taken from two sources. The first source is the list compiled by Mr. R. C. Eldridge of Niagara Falls[1] in his analysis of the vocabularies of 250 different articles taken from newspapers, and the second is the "Foundation vocabulary" of the 1000 commonest words, compiled by Dr. Leonard P. Ayres, and used as the basis for his spelling scale.[2]

[1] Eldridge, R. C. *Six Thousand Common English Words*, Niagara Falls, New York, 1911.

[2] Ayres, Leonard P. *A Measuring Scale for Ability in Spelling*, Russell Sage Foundation, New York, 1915.

Since the thought of all the paragraphs is kept of uniform difficulty and simple, it has been an easy matter to make the style conform to that of ordinary newspaper English. The style is not so simple as that of the first grade reading book, nor is it as hard as much of the material found in text books. The standard of the newspaper story has been, so far as possible, maintained throughout.

It should be noted, however, that while style and vocabulary are like those of the ordinary newspaper, the type of reading required by the test is of a different nature from the newspaper requirement. Newspaper reading calls for the ability to read rapidly but it does not usually require exact attention. The scale calls for reading with careful attention to details, and the pupil is required to read each paragraph well enough to follow its instructions correctly. It is a test of careful reading.

Memory Span

Since these scales do not employ tests in reproduction the chief memory obstacle to measuring reading is avoided. The attempt has been made to do two things: first, to reduce the number of ideas which the child must hold in mind to so few that any normal child is able to carry them. To do this all the paragraphs have been re-written until the separate ideas presented are gathered in, and made subsidiary to one central thought. Each separate idea has been made so much a part of the central concept that the child is not aware of any separation between them.

The instructions are sufficiently definite so that the child cannot follow them correctly without reading the entire paragraph, but they are knit together closely enough to seem like a single direction to the child.

The second rule followed in the construction of the scale is to keep the memory requirement uniform through all the paragraphs. So far as possible, no heavy memory loads have been permitted to enter; and where experiment indicated that the memory element complicated the reading difficulty the paragraph has either been removed or re-written.

Planned for Classroom Use

The third important characteristic of Picture Supplement Scale 1 is that it is simple to administer and easy to score. While there is need for much individual testing, by far the greater part of measuring school results must be done by groups or classes if it is to be done at all. It was therefore decided to make the new silent reading scale in such a way that an entire class could be tested at one time. To make this feasible it was also necessary to have the test short so that it could easily be carried through in part of one period, and to have the method of scoring quick, easy, and accurate.

A time limit of exactly five minutes has been established for testing. The time allowed is long enough to prevent accidental delays in starting, etc., from having too great weight in the final score. It is short enough to avoid an accumulation of undistributed perfect scores, due to children finishing before the time is up.

The advantage of using the picture device is obvious when the problem of scoring is under consideration. The teacher who has read the paragraphs of the scale knows exactly what is required of the child. She does not have to consult an answer list or carry through complicated mental calculations to see if the child is right; she has merely to glance at the picture. The matter of finding the score becomes the simple process of counting the number of pictures correctly marked.

If scales are to be used in classroom work, they must be inexpensive. It was decided to cut out every unnecessary item of cost. Illustrations are small and easily reproduced. Sheets are used instead of pamphlets; first, to avoid the costs of folding and stapling, and, second, because they have proved considerably more convenient in the classroom, since the children cannot start at the wrong place, or turn over two pages at a time, and the teacher does not have to turn pages in scoring.

Another, and important, feature which seemed desirable was that the new scale should be made available in several forms, so that tests could be repeated at frequent intervals; and that after the first results were secured, additional scales could be prepared from time to time as the need for them became evident. The simplicity of the materials and of the method adopted has made this possible. By the time this book is published, Picture Supplement Scales 2, 3, and 4 will probably be ready for distribution. These scales are of the same type and

difficulty as PS-1. They may be used interchangeably; and scores in one may be directly compared with scores in another.

Grade Scores Assigned Equivalent Grade Values

The fourth outstanding characteristic of Picture Supplement Scale 1 is that the results of testing are secured in the form of distributions of scores for each separate grade, from the third to the eighth inclusive. These distributions form the basis for grade scales in which each number of paragraphs correctly read is assigned a credit or value which indicates where it stands along the base line of its grade distribution. That is, the scales show for each grade whether a given score is the worst score commonly found among 100 typical children of that grade, or the best score, or the middle score, or what other position it holds with reference to all the scores which are commonly found for children in the grade in question. It is possible, therefore, in testing with Picture Supplement Scale 1, for each child to be judged by a jury of his peers; his ability is measured in terms of its relation to the known abilities of other children who are approximately of the same degree of maturity, and have received approximately the same amounts of training as himself.

Summary

1. Every scale for measuring silent reading is an attempt to answer the teacher's question, "How much

did the child get out of what he read?" by having him do something which he can do correctly only if he understands what he reads. The three ways in which this has usually been attempted are to have the child read and reproduce; or read and answer questions; or read and follow directions.

2. Four limitations of existing tests and scales are recognized. These are, first, that they are not all genuine measures of reading ability; second, that where they consist of a series of tasks, these tasks are not consistent in character; third, that they are hard to administer and difficult to score; and fourth, that they do not always furnish data whereby the achievements of an individual child can be compared with the achievements of other children in his own grade.

3. The new scale for measuring silent reading, Picture Supplement Scale 1, is designed to measure reading by strictly utilitarian standards. The attempt has been to devise a test in which the child can readily succeed if he reads well enough to grasp the important thought in each section, and in which he cannot succeed at all unless he does comprehend each important thought. The amount of this kind of reading which the child can do in five minutes is the variable which is to be measured.

4. The scale has four outstanding characteristics. The first is that it makes a definite attempt to measure a single ability, which is the ability to read silently a single type of material, at a constant level of difficulty, in a fixed period of time. It measures the

amount of reading of a practically useful nature which the child can do in five minutes.

5. The second outstanding feature of the new scale is that a careful attempt has been made to discover the controlling factors in silent reading. Some 25 such factors have been identified. One, the child's rate of reading, has been adopted as the variable to be measured; and the remaining 24 factors have been, in so far as possible, held constant. It is believed that by following this method, a test has been prepared in which every task presents the same type of reading difficulty as every other, and for which the scores represent comparative amounts of one single sort of reading ability.

6. The third outstanding feature is that the test is planned for classroom use. It can be given to large numbers of pupils simultaneously. It requires five minutes for actual testing; and can be scored accurately, rapidly, and easily. The cost of printing has been kept low; and companion editions can be prepared as need arises. Three such alternate editions have already been prepared as Picture Supplement Scales 2, 3, and 4.

7. The fourth outstanding feature is that grade scores have been turned into equivalent scale values for those grades. This makes it possible, in testing with Picture Supplement Scale 1, to measure the ability of each child in terms of its relation to the known abilities of other children who are approximately of the same degree of maturity, and have received approximately the same amounts of training.

CHAPTER III

PRELIMINARY EXPERIMENTS IN MEASURING SILENT READING

During the experimental work which produced the scale that has been described in the previous chapter, several other tests and scales were developed and tried out in different city school systems. In all, six scales were printed and slightly more than 10,000 copies of them were used by children in the public schools of nine cities. One of these was the scale designated as PS-1 which has already been considered. The remaining five will be described briefly in this chapter in the hope that the experiences recorded as to their nature and the results of applying them may be of use to students of educational measurement.

These five scales were all of the same general plan as PS-1; that is, they were all on sheets of paper 11 inches wide and 19 inches long, divided into five columns of four divisions each. All the scales consisted of pictures and paragraphs of instructions about the pictures. On the pages that follow brief accounts will be given of each of the five preliminary scales. No attempt is made to reproduce them in full, with pictures, method of scoring, etc., but the text of each

is given, in the appendix at the end of this volume, in the hope that it may be of suggestive value.

Two Hearing-Reading Scales

The first two scales to be developed were given the office designations of HR-1 and HR-2. These were the Hearing-Reading scales. The Hearing-Reading scales were printed on both sides. On the front side were 20 pictures. The teacher read aloud instructions for marking each picture with a pencil, and the children, having listened to the instructions, proceeded to follow them. After all the pictures on the front of the sheet had been so marked, in accordance with instructions to which the children had listened, the sheets were turned over. On the back was a similar set of pictures, but in this case the instructions were printed directly beneath each one. Instead of listening to the teacher reading aloud, the children read the paragraphs for themselves and, as they finished silently reading each paragraph, they followed the instructions it gave by marking the picture above it with their pencils.

Both the paragraphs read aloud by the teacher and the printed paragraphs that the children read to themselves were graded in ascending difficulty of vocabulary from very simple ones at the beginning to extremely difficult ones toward the end of each scale. Each paragraph on one scale corresponded to another of the same difficulty on its companion scale. The underlying idea of these hearing-reading scales was to use them first to secure a record of the ability

of the child to follow instructions received through hearing them, and then to secure a second record of his ability to follow another set of instructions, of equal difficulty, when he had to get their meaning by reading them himself. Both sets of records having been secured, they were to be compared in order to find out how nearly the child, when reading, could equal the record that he made when he listened and did not have to read.

The Hearing-Reading plan was laid aside for two reasons. The first reason was that it was found impossible to control the administration of the oral testing by the teacher, so that it would be constant for all children. Some teachers read fast and others slowly; some enunciated clearly and others spoke with such a marked accent that the children could understand them only with difficulty. Moreover, since the paragraphs of the hearing-reading scales increased in vocabulary difficulty, there were some teachers who were unable to pronounce all the words, and so did not succeed in reading them properly aloud.

The second problem encountered in the hearing-reading scales was the difficulty of interpreting the scores. In general children who are just beginning to read do better when hearing instructions than when reading them; but after the early stages of reading have been passed, the relationship is usually reversed; so that in the upper grades, where children have acquired facility in reading, they are likely to make a better score with printed instructions than with

those which are given to them verbally. It was evident from the results of preliminary testing that the hearing-reading relationship gives promise of valuable material for the diagnosis of reading ability; but it was also clear that much more careful and extended experimenting would be necessary, in order to secure valid results which could be readily interpreted, than was possible in the present study.

There has not been time to make any hearing-reading experiments with the standard silent reading scales now in use, but apparently this might yield valuable results. The instructions or stories of the different scales might be read aloud by the examiner, and the children asked to listen, in order to reproduce, answer questions, or obey orders, according to what the tests call for. While the results of the present experiments are too few on which to base valid conclusions, there is some reason to expect that for most readers lower scores will be made when the material is read aloud to them than when they read it themselves. This hearing score seems to be progressively lower as the material includes non-reading difficulties such as puzzles, catches, mathematics, abstract reasoning, memory tests, and the like. The less the selections measure pure reading ability, the lower the hearing score seems to be.

The oral use of silent reading scales also suggests a method for individual diagnosis. If the hearing score of a pupil is exceptionally low as compared with that of other children with the same amounts of training, it may indicate that the pupil's difficulty is not pri-

marily that of reading, but has to do perhaps with unfamiliarity with English, inability to pay attention, poor memory, deficient vocabulary, and the like. The experiments briefly reported here showed that, when no time limit is set for silent reading, pupils having high scores in the hearing test do even better in the reading test, but pupils with low scores in the hearing test seem to follow no particular rule as to their achievements in reading. There was a wider distribution in the silent reading scores than in the hearing scores.

Scales of Increasing Difficulty

The original hearing-reading scales were divided into 20 steps of steadily increasing difficulty. Thought, style, memory span, and task were all maintained at a constant level, and the vocabulary was carefully varied. After the hearing-reading plan had been laid aside, the attempt was made to use these scales as measures of reading difficulty.

A preliminary question had already been raised as to whether vocabulary difficulty could properly be increased beyond a moderate level, since, in constructing the paragraphs it had been found that it is practically impossible to maintain a level of natural unstilted English prose while at the same time using harder and harder words. Hard words are not used in ordinary writing; and their introduction at once tends to make the writing artificial. Normal English consists of words so simple that the ordinary adult does not have to look them up in the dictionary.

The two sets of 20 paragraphs of increasing vocabulary difficulties were given as straight difficulty tests to a large number of children. No time limit was set, other than that of the ordinary classroom period, and it was assumed that the children starting at the beginning of the test with the easiest paragraph would work their way through harder and harder paragraphs until the words became so long and difficult that the children would not be able to understand what they meant. It was expected that the hardest paragraph a child could succeed in doing would mark the upper limit of his ability.

What the experiments disclosed was that, if the thought is simple and the style is simple, hard words will not usually prevent children from getting the gist of a paragraph. They merely increase the amount of time required. When children meet hard words they jump them; grasp the easy words which they do understand; and with the easy words as key words, piece out the meaning of the others. The result was that when these scales of increasing vocabulary difficulty were tried out in the classroom, and no time limit was set, there were practically no low marks. Even in the third grade most of the children made perfect or high scores. The children probably varied greatly in reading ability, but there was little variation in their scores.

The experiment was then tried of setting a time limit. The children were allowed to work for five minutes, and it was desired to find how high they could go in that time. The distributions of scores

under this new method were more satisfactory, in that they were widely distributed between 0 and 20. Close observation of children taking the test, however, disclosed the fact that they were working at markedly different rates of speed. Some children read as fast as they could, skipped some paragraphs entirely, half-read others, but made at least an attempt on every one of the 20 paragraphs before the five minutes were up. Other children worked fast until they came to a problem which puzzled them, and spent the remainder of the period unsuccessfully trying to solve it. Still other children, working at much slower rates of speed, succeeded in doing every problem they reached, and showed clear evidence of being able to keep on ascending when they were suddenly cut short. It was clear to the examiners that the scores of these children were not truly comparable, since ability to do difficult work and ability to work fast were indiscriminately mixed.

These experiments led the examiners to believe that in devising scales for difficulty, special consideration must be given to the element of time. If the test is to be given without a time limit, evidence must be presented to prove that the amount of time allowed has no effect upon the score. If time is found to be an important controlling factor, a definite time limit must be set, or a definite record of each pupil's rate of working must be made, not for the test as a whole, but for each separate level of difficulty within the test.

Two Continuous Narrative Scales

Two Continuous Narrative scales, CN-1 and CN-2, were prepared on a different plan from any of the others. Each scale consisted of a short and interesting story which was divided into 20 sections of equal length and equal difficulty. These sections were then so arranged that the location of each one could only be found by reading the section preceding it. The child was allowed to read for five minutes.

Scale CN-1, like the other scales of this series, was a single sheet of paper 11 inches wide and 19 inches long. The sheet contained five columns. Each column was divided into four sections, and in each of these sections there were a picture and a paragraph about the picture. The whole 20 paragraphs told the story of how a little Prince learned to like books. This story did not run consecutively from paragraph to paragraph, but instead the paragraphs were scattered among the different sections, and each, while carrying the child one step forward in the plot, told him how to find the new paragraph where he could read what happened next. The material printed below gives the first three paragraphs of the story in the order in which they came if read correctly by the child.

The Prince's Book

Once upon a time there was a lazy little Prince. He knew how to read, but he did not like to do it. His Father the King was naturally very much worried, and finally he called Pen and Paper to help him. "They will know how to make my little son like books," he thought, and he sent a messenger to call them. Now Pen and Paper were the first people who

tried to help the Prince to like books, so you must find their picture, which is just below this, and write a figure 1 beside it with your pencil. Then go on reading to learn what they said to the King.

"The Prince will like books," said Pen and Paper, "if we write one for him." "Write it, and send it by the Postman," said the King. As the Postman was the second person to help with the Prince's book find his picture in the last column and write 2 beside it. Then go on reading to learn what he did about the book.

"Dear me!" said the Postman, "This is a heavy book. It is full of stories for the little Prince I must run to the castle and give it to him." If you will look in the third column you will find a picture of the Prince reading his book, and as the book was the third one to help him, write 3 beside it, and read what happened next.

Scale CN-2 was on exactly the same general plan as CN-1. In this case, however, there was one-seventh more material in each paragraph, and the thought was somewhat harder. The full text of both continuous narrative scales is given in the appendix. It will be noted that in the Continuous Narrative scales the paragraphs uniformly carry two instruction thoughts apiece. One of these tells the child where to find the paragraph that follows in the sequence of the story. The other tells him what number he is to write beside the picture of the paragraph when he has located it. Classroom experimentation seems to show that to the child these numbers are a necessary part of the story. In CN-1 he writes them down in order to keep track of how many people helped in teaching the little Prince to like books, and in CN-2 he is asked to number the pieces of evidence in the order in

which they were presented at the government trial where John testified against the band of spies. The children do not know it, but the fact is that in each scale these numbers run from 1 to 20 and indicate how many paragraphs the child has read up to that point. The result is that the child leaves a record behind him which tells the teacher where he went and what he did. If he has read and marked the story correctly, the highest number written is his score on the test. The test can be correctly scored almost as fast as the teacher is able to look down the columns.

Scales CN-1 and CN-2 have been tested in all grades from the second through the eighth, and the records compared with those made by the same children on PS-1. In general it may be said that the narrative scales require a less exactly careful type of reading than do the picture supplement scales. Pupils make far more diversified scores with the narrative scales, and, in general, higher ones. Apparently the two types of scales measure products which differ considerably from each other.

The Continuous Narrative scale, as exemplified by CN-1 and CN-2, was laid aside in the present experiments because it seemed to present more problems, and to measure a type of reading somewhat less important for classroom use than did the Picture Supplement scales. The continuous narrative, however, seems to be a device which is capable of development into a valuable instrument for measuring reading. The essentials for making such a scale are, first, to procure a large number of pictures from which

to select a complete set of illustrations which may be combined into an interesting and exciting narrative, and which are sufficiently different so that a description of one will not be mistaken for that of another. The other requirements are ample time for writing, experimenting, and re-writing, funds for printing large numbers of trial copies for experimental purposes, and generous amounts of care and patience.

A Difficult Picture Supplement Scale

At the same time that PS-1 was being constructed a companion scale was prepared for testing in the same way at a somewhat higher level of vocabulary difficulty. The following paragraph is typical of those used:

> "To shake the poise of this unpleasantly supercilious butler outline a large boulder immediately in his course. This will almost inevitably cause him to stumble and be precipitated headlong; but in order to insure his demoralization draw still another a short distance further in front of him."

In making this scale it was found that slight increases in word length had very little effect upon the pupil's ability to get the meaning quickly. When considerably longer words were introduced the difficulty was increased, but the material became stilted and unnatural, so that care had to be exercised to prevent the paragraphs from becoming a type of English which was not truly representative of the ordinary material which people are called upon to read. The Difficult Picture Supplement scale has been tested in all grades from the third through the eighth, and when

taken by the same children it was found to be considerably more difficult than PS-1, CN-1, or CN-2.

Principles Involved

In the course of the investigation which led to the production of the new scale, Picture Supplement Scale 1, several principles were recognized which apparently are fundamental to the construction of scales for measuring ability in silent reading. The five following chapters will be devoted to considering the nature of these fundamental governing principles, how they operate, and what is involved in applying them to the measurement of reading. The first of these principles is that discussed in Chapter IV, which deals with the Law of the Single Variable.

Summary

1. In the course of the investigations recorded in this book, six different scales for measuring silent reading have been completed, printed, and tried out in 23 school systems. The experimental copies used in this way have been more than 10,000 in number.

2. Of these six scales, two have been designated Hearing-Reading Scales; two others Continuous Narrative Scales, and the remaining two Picture Supplement Scales.

3. The Hearing-Reading scales are so devised that the child, in marking one side of the scale, is following instructions which he hears from the lips of the teacher, while in marking the other side of the scale he is following instructions which he reads.

4. The Continuous Narrative scales are ones in which the child numbers, as he reads them, the consecutive paragraphs of a story. These paragraphs are scattered about on the page, and in order to find them he must fully understand what he reads.

5. The Picture Supplement scales consist of pictures and paragraphs so arranged that the pupil draws with his pencil a line or lines to supplement the picture. His ability to do this in accordance with the printed instructions reflects the rapidity and accuracy with which he can read.

6. The scale chosen for full development in connection with the present study is one of the Picture Supplement scales, which has been given the office designation of PS-1.

7. In the course of the investigation here described, several principles were recognized as fundamental to the measurement of silent reading. These principles will be discussed in the five following chapters. The first principle to be considered is treated in Chapter IV under the heading, "The Law of the Single Variable."

CHAPTER IV

THE LAW OF THE SINGLE VARIABLE

In the ordinary classroom test in arithmetic, the children are given an examination consisting of ten problems. At the end of the period, the papers are corrected and marked. A pupil who has done the first three examples correctly is given a mark of 30; one who has six right is given a mark of 60; and the bright pupil who correctly solved all the problems before the end of the period receives a mark of 100.

Such marks are ordinarily accepted by school people at their face value as measuring the relative accomplishments of the pupils. The mark of 60 is taken as representing twice as good a performance as that of 30, and the one of 100 is accepted as being twice as good as one of 50. In recent years the newer scientific movement in education has produced evidence showing that such marks as these cannot be accepted as trustworthy measures of the relative abilities and achievements of the pupils, and indeed that they are often seriously deceptive.

The reason why the marks received by the pupils in the arithmetic test usually do not accurately record their abilities and the value of their achievements is that they do not measure different amounts

of the same thing. In the first place, the different examples in such a test are frequently of widely varying difficulty so that it may well be that the pupil who got only three right may have had almost as much arithmetical skill as the one who got six correct. If he had had a little more time, after having solved the third example, he might perhaps have done five or six correctly by the end of the period.

Again, the mark of 100 does not correctly reflect the skill of the brightest pupil, as compared with the others, for he finished before the end of the period and is penalized because there was no more material for him to work on. If there had been, he might have finished 13 or 14 examples.

A further shortcoming of such a test is found in the fact that the work of the pupils was not only on material of varying degrees of difficulty and worked on for different amounts of time, but the quality of work done by the pupils was at different levels of excellence. Some worked much more neatly than others and arranged their material more intelligently. Some missed the correct answers because of minute errors, such as slips in copying, while others submitted answers that were not only incorrect but absurd and obviously impossible.

The marks resulting from such a test are crude measures of the accomplishments resulting from undertaking tasks composed of units of differing degrees of difficulty, worked at for varying amounts of time, and producing solutions, or attempts at solutions, of varying degrees of quality. The marks do

not measure amounts of any one thing. They measure conglomerates composed of achievements conditioned by the three factors of quality of product, difficulty of task, and time consumed.

Consistency

It is to remedy these shortcomings of our marking systems that the scientific movement in education has devised its tests and scales. These measuring instruments apply to classroom processes and products the fundamental law of physical measurement which is that the thing to be measured must possess the quality of consistency. It must remain constant while it is being measured. This is a fundamental necessity for logical thinking about measuring, counting, or enumerating. The things counted must all be of the same category. The thing measured must be constant in its character or composition so that one unit of it will be equal to any other unit of it.

The Single Variable

The process by which the essential characteristic of consistency is obtained in educational measurements is the one used in physical measurements. It consists of distinguishing the possible controlling, varying factors; devising means for holding them all constant save one; and measuring that one. This is the law of the single variable.

The importance of the law is readily seen in the measurement of handwriting. Any close comparison

of the merit of achievement must be in terms of only one variable, which, in this case, will be the quality of handwriting. If the handwriting of one pupil is to be considered as representing a more meritorious performance than that of another pupil, it must be shown that the material written was the same in both cases, that it was written under the same external conditions, that the material was equally familiar to both children, that they both wrote for the same amount of time, and succeeded in writing the same amount of material within that time. If, under these conditions, the quality of one handwriting is better than that of the other, it may truly be said that one represents a better achievement than the other.

The Courtis Tests in Arithmetic

The fundamental difference in method between the non-scientific test which undertakes to measure an uncontrolled composite of different variables, and the scientific test which carefully restricts its measuring to a single variable, while holding the others constant, is best illustrated by citing the arithmetic tests devised by Mr. Stuart A. Courtis, of Detroit. This pioneer student of educational measurement early recognized that in testing children for arithmetical ability several different factors would exercise a controlling influence on the results, unless special pains were taken to restrict the measurement to one factor among them, while holding the others constant.

He did this by devising a form of test consisting of a large number of problems of the same sort and of

equal difficulty, printed on prepared sheets. All of the children of a class begin to work at the test simultaneously, and continue for exactly the same period of time. The test contains more examples than any of them can complete within the time limit. The resulting scores measure the comparative abilities of the children in the particular phase of number work that is to be tested.

The results are trustworthy indicators of the varying abilities of the different children, because they are in terms of greater or smaller amounts of the same thing; that is, they are in terms of units of the same kind within a given time limit. The conditions of the test are equal for all the children; the time is uniform; the work is held at a constant difficulty; and the element of quality is substantially eliminated by having the examples printed so that pupils have only to write the answers in the appropriate indicated places. What Mr. Courtis has done in devising his test is to observe scrupulously the law of the single variable.

"Other Things Being Equal"

When children first attend school and begin to study arithmetic the teacher impresses upon them the principle that they can count, add, subtract, multiply, or divide only in units of the same category. When they get a little older, they learn that measurements can be made only in unvarying and clearly defined units. A little later on, they begin to apply the law of the single variable to different sorts of comparisons, outside of the realm of exact measurements.

As a general rule, they do this without having carefully formulated the law itself; they recognize the necessity for respecting it by saying that "other things being equal" such and such a result will follow. In saying this, they vaguely realize that in making comparisons of things where various interdependent elements enter, the other factors must be kept constant, and one single element measured.

This principle has been carefully followed in devising the present test for the measurement of silent reading. The test has recognized the interdependent controlling factors, selected one as the variable, and contrived, in so far as possible, to keep the others constant. It has done this by keeping the difficulty of the task and the quality of the work as nearly constant as possible, and measuring the amount of achievement attained by the pupils in a given period of time.

Summary

1. The traditional classroom examination consists of tasks of varying units of difficulty, worked at for varying amounts of time, and producing results of varying degrees of quality. The examination mark does not measure any one thing. It measures a conglomerate of achievements conditioned by the three factors of quality of product, difficulty of task, and time consumed.

2. It is a fundamental law of measurement that the thing being measured must be consistent in character or composition so that each unit of it will be equal to every other unit.

3. This essential characteristic of consistency may be obtained in educational measurements by distinguishing the possible controlling varying factors, devising means for holding all of them constant save one, and measuring that one. This is the law of the single variable.

4. The application of this law is well illustrated by the work of Mr. Stuart A. Courtis of Detroit, in the measurement of ability in arithmetic. In the Courtis tests, the conditions of the test are equal for all the children, the time is uniform, the work is held at a constant level of difficulty, the element of quality is controlled by the form in which the test is given, and the variable that is measured is the amount done in a given length of time.

5. The law of the single variable is a principle of measurement taught in the earliest school years, and increasingly recognized in the comparative judgments of everyday life.

CHAPTER V

THREE TYPES OF SCALES

The law of the single variable is to the effect that in careful comparative measurements every controlling varying factor must be identified, one must be chosen as the variable to be measured, and all the others must be held constant.

The factors which affect results of different sorts of testing are so numerous that no one person could identify them all. Even for a single case the powerful controlling influences will fall in lists of 20 or 30, and each one of these has subdivisions within itself. When the attempt is made to classify them, however, so that all the controlling factors of one sort are grouped together, it will be found that no matter what subject is under consideration, or how many controlling varying factors may have been identified for it, each of these factors may be classified as falling into one of three main groups. It is either a variable of quality of product, or of difficulty reached, or of amount done.

Marksmanship, A Measure of Quality of Product

Quality, difficulty, and amount are the three variables. One of them must always be measured; the

other two must always be controlled. Measurements for these three variables may be illustrated by citing examples from athletic contests. The measurement of quality of product is illustrated by the contest in marksmanship. When people are shooting at targets, the range is fixed, the targets are of uniform size and shape, and the centers or bull's-eyes towards which the marksmen aim are equal in color and diameter. These are elements which affect the difficulty of shooting; and they are made as nearly as possible equal and constant for all those who shoot. Since each contestant is given an equal number of trials, the amount which he shoots is fixed. One element, however, is not fixed, but is allowed to vary; and the records of its variations form the score of the contest. This variable, the different degrees of which are noted, is the quality of the marksmanship. It has no definite limits of right or wrong; it ranges all the way from shots so wild that they are barely distinguishable as attempts to hit the target to shots which land exactly in the center of the bull's-eye. Marksmanship is a measure of quality.

Bowling is another measure of quality of product. Difficulty is kept constant by establishing standards for distance, surface of alley, set up of pins, and so on; which are uniform for all players. The amount is controlled by limiting the number of attempts to two; and the quality of the performance is measured by the proportion of pins knocked down in these two trials.

Jumping, A Measure of Difficulty Reached

The high jump, however, is not a measure of quality of product; it is a measure of difficulty reached. Quality is kept at a constant level; it is "good enough to clear the bar," and that degree of quality is set as a passing mark which every contestant must meet if his performance is to score. The contestants are, as a rule, allowed three attempts at each height; so that the numbers of chances for success are equal. The element which is allowed to vary is the height the contestant can jump, and, as each height is successfully cleared, another and more difficult height is set. This sort of contest, where quality is set, the amount or number of trials is uniform, and the score marks the difficulty of the hardest task successfully completed, is a measure of difficulty reached.

Racing, A Measure of Time Consumed or Amount Done

The third sort of athletic contest is illustrated by the race. In most forms of racing the quality element plays a minor part. There are usually two minimum requirements for quality. The contestant must run well enough to finish the course; and he must refrain from interfering with or fouling his competitors. Quality in excess of these requirements has no effect upon the score. The difficulty of the race is made equal for all the contestants, by having them cover the same course. With quality and difficulty held constant, amount done is left as the third element which must be treated as the variable.

The variable of amount is handled in two ways. In some races the contestants are allowed to ride or run for a definite period of time—as in the case of the six day bicycle race—and the resulting scores are in terms of the amount of ground covered. In other races a definite distance is set, and the scores show how much time was required to cover that distance. Time and amount are complementary terms. Time implies amount and amount implies time. In the three-fold classification of variables which can be measured, time and amount have been treated as one group, under the word "amount," because they are essential to each other. The question "How much can be done?" demands at once a statement of the time allowed for doing it; and the question "How long will it take?" depends upon how much there is to do. Time and amount must always be considered together.

Shooting, jumping, and racing have been cited as typical examples of the three fundamental sorts of measurement; but it would be possible to expand such a list to include practically every sort of athletic activity in which comparative achievements are reflected by carefully recorded quantitative scores. It will be found that nearly every careful quantitative measurement we make seeks to measure one of the three fundamental variables—quality of product, difficulty reached, or amount done—and in order to measure that one variable, seeks to control the other two.

Three Types of Scales

When the classroom teacher wishes to judge the work of one of her pupils, she asks herself three questions about him. She wants to know "How well can he do?" "How hard work can he do?" and "How fast can he do it?" and the answers to her three questions furnish the basis on which she decides whether to keep him, to promote him, or to send him back to the grade below.

These three questions, how well, how hard, how fast, represent the teacher's attempts to measure the three fundamental factors of quality, difficulty, and time or amount. The educational tests and scales which have been devised during the past ten years are attempts to help her answer those questions; and each of them seeks to measure some one of those same three fundamental factors. While the principle of the single variable has not always been fully understood or closely followed, in general it may be said that the standard educational measurements fall into three clearly defined groups, according to which of the three fundamental factors they have chosen as the variable they seek to measure. They are tests and scales for quality of product, for difficulty reached, and for amount done. Moreover, it is probably true that new scales as they are developed in the future must inevitably belong to one of these three groups; and the student of educational measurement who plans to devise a scale must seriously consider which of the three variables he will attempt

to measure, and which of the three types of scales the one he presents must therefore be.

Summary

1. The innumerable factors which influence the results of testing may be classified into three distinct fundamental groups. They are variables of quality, of difficulty, or of amount.

2. The measurement of quality is illustrated by contests in marksmanship or in bowling. In these contests the difficulty of task and the time allowed for doing it are maintained as constants, and the variable measured is the quality of the performance.

3. The measure for difficulty is seen in the high jump. There quality and time are constants, and the variable is the difficulty of the hardest task successfully done.

4. The measure for amount is seen in the race, where quality is but slightly operative, difficulty is constant, and the variable measured is either the amount done in a given time, or the time required to do a given amount.

5. Time and amount are complementary terms, each of which depends for its meaning upon the other. In the threefold classification of variables, the term *amount* is to be considered as carrying with it its companion term *time*.

6. Educational measurements are attempts to answer the three fundamental classroom questions: "How well can he do?" "How hard work can he do?" and "How fast can he do it?" Each seeks to mea-

sure one of the three fundamental factors, and, according to which it selects, it may be classified as a test or scale for quality, for difficulty, or for amount.

7. The student of educational measurement who plans to devise a scale for ability in any school subject must consider, first, which of the three variables he will attempt to measure; and second, having chosen that variable and thereby fixed the type of scale which must be employed, what are the implications as to the methods he must follow.

CHAPTER VI

SCALES FOR QUALITY OF PRODUCT

There is a group of school subjects in which the ordinary classroom work results in tangible recorded products of varying qualities. Among such products are handwriting, freehand lettering for mechanical drawing, drawing, and, less definitely, English composition. In these subjects the question the teacher asks is "How well can the child do?" and the measuring device which must be used to answer her question is that of the scale for quality of product.

From the point of view of educational measurement the outstanding characteristic of handwriting is that the thing itself is there to be measured and that it exists in samples of differing degrees of quality. Some of the handwritings are poor, when the ages and grades of the children are taken into account, while others are fairly good, and still others may be considered excellent. The problem of the measurer is to gage or determine the relative degree of goodness of the actual sample lying on the table before him.

The sample itself has physical properties. If the writing were perfect its lines and letters would be regular, its slants uniform, and its spacing equal. In proportion as the symbols on the paper depart from

these known standards, the goodness of the writing is diminished. These diminutions, moreover, take place by infinite varieties of combinations and permutations in a perfectly continuous series down to the point where the writing is so bad as hardly to be handwriting at all. There is no right or wrong in the quality of handwriting. It simply ranges from less good to more good through a continuous series of degrees of quality.

The same observations may be made with respect to freehand lettering. Here again the product is tangible and objective, and the quality of a given sample may be gaged by its deviations from those set characteristics which would constitute a perfect sample of the sort of lettering in question.

With a less degree of definiteness similar comments may be made with regard to samples of drawing. The products of work in manual training have the same characteristics and belong in this same group from the viewpoint of the measurer. Still another type of product in this group is that of composition. As in the case of the handwriting, lettering, and drawing, the actual product is available for examination and comparison. Moreover, the samples exist in varying degrees of quality which range from poor to good in unbroken series.

The classroom products that have been considered, writing, lettering, drawing, and composition, are measurable by scales for quality. Their common characteristics make it possible to construct such scales, and in the case of each of them, there are in

existence educational measuring instruments of the quality type which have demonstrated their value and validity.

Measuring scales for quality of product were the first type to be developed. The reason for this is that the pioneer students, entering a field that at best presented serious obstacles, undertook at the outset the measurement of products that existed in the shape of tangible records that could be examined and compared as many times and by as many methods as might be necessary.

In the modern movement not only was the first scale one devised by Professor E. L. Thorndike, for measuring quality of handwriting, but other quality scales for lettering, drawing, and composition followed shortly. Again, it was Professor Thorndike who developed the first scale for the measurement of drawing, and published it in 1913.[1] During the same year, Dr. H. O. Rugg developed his scale for the measurement of freehand lettering which he published two years later.[2] The pioneer work in the measurement of quality in English composition was produced even earlier, and published by Dr. Milo B. Hillegas in 1912.[3]

[1] Thorndike, E. L. The measurement of achievement in drawing. *Teachers College Record*, 14: Nov., 1913, pp. 345–383.

[2] Rugg, H. O. *A scale for measuring freehand lettering.* University of Chicago, Chicago, Illinois.

[3] Hillegas, M. B. A scale for the measurement of quality in English composition by young people. *Teachers College Record*, 13: September, 1912, pp. 331–384.

In the scales under discussion, quality is chosen as the variable which is to be measured. In accordance with the law of the single variable, then, the remaining two factors of difficulty and amount, with all their numerous subsidiary elements, must either be excluded from the test or so rigidly controlled that they are restrained from influencing the results.

What this means in the case of scales for quality of handwriting has already been touched upon in Chapter IV, in the discussion of the law of the single variable. It was pointed out that accurate comparisons between the achievements of children in handwriting can be made only where the writing has been produced under certain carefully controlled conditions. The testing must take place under similar external conditions of seating, lighting, and so on. Children must be furnished with equally good writing implements. The material to be written must be the same for all the children, and equally familiar to all of them. They must write for the same length of time; and must have finished equal amounts within that time. If it can be shown that these conditions have been met, comparisons between the qualities of the samples of handwriting produced under them can be regarded as true comparisons of achievement; for the comparisons based on quality are then unadulterated by intruding factors of difficulty or time.

Similar observations may be made for any subject in which the quality of the product is to be measured.

When quality is chosen as the variable, every remaining factor which contributes to the difficulty of test or the time required to take it, must be identified; and, having been identified, must either be excluded or controlled.

Reading Not Measured By Scales For Quality

Reading is a classroom activity which does not readily lend itself to measurement by means of scales of quality. One reason for this is that it does not result in a tangible objective product which can be scrutinized and measured. Another reason is that quality in reading is an elusive thing which varies not only with different people but with the same person from moment to moment as he reads. Dr. C. T. Gray records, in his monograph on Types of Reading Ability,[1] experiments in studying breathing and perception span in oral reading by having the subject read aloud into the receiver of a dictaphone. The wax records thus made served as the basis for close analytical study. It might be feasible to use a similar device in securing records of oral reading for the purpose of measuring expression, pronunciation, and emphasis; but the great difficulties in procedure of this sort of testing prevent its use for the present in public schools.

There is more reason, perhaps, to wish for some

[1] Gray, Clarence Truman. *Types of reading ability as exhibited through tests and laboratory experiments;* an investigation subsidized by the General Education Board. University of Chicago Press, 1917. See pages 70, 75, 127.

method of measuring the richness of mental imagery, the variety of neurone connections, which are stimulated by the printed page. The reason that such qualitative measurements cannot easily be made is that there are as many different things which children get out of reading as there are children and as there are times that children read. Because of the infinite combinations of neurone connections in different minds the reading process awakens in consciousness thoughts and memories of the most varied character. In re-reading even a simple passage the same individual gets new meanings from the page, new and different mental reactions from the same stimuli.

Moreover, outside of the psychologist's laboratory the shades of quality of reading are not of great import. For practical purposes the problem of measuring reading involves finding out how rapidly the subject reads the material with a sufficient degree of comprehension to get from it the essentials of its meaning. As the material increases in difficulty the ability to read it is rarer; but the important question is not "How full and varied is the meaning the reader draws from it?" but rather "Is he able to grasp the gist of the material?" The quality element is reduced to the very simple one of "Well enough to get the essential thought."

Summary

1. Ability in such subjects as writing, drawing, and composition, in which the products are tangible rec-

ords of varying degrees of quality, is best measured by scales for quality of product.

2. The first of the modern scales for measuring classroom products was Professor Edward L. Thorndike's scale for measuring the quality of handwriting. This was a scale for quality of product.

3. Other scales for quality of product quickly followed. Among these were Dr. Rugg's scale for measuring freehand lettering in mechanical drawing, Dr. Thorndike's scale for measuring the quality of children's drawing, and Dr. Hillegas' scale for the measurement of quality in English composition.

4. Scales for quality of product select quality as the variable which is to be measured. In accordance with the law of the single variable, they must rigidly control the two remaining variables of difficulty and amount, and their subsidiary elements, in order that these variables may be prevented from influencing the results.

5. Reading does not readily lend itself to measurement by scales for quality of product. One reason for this is that it does not directly result in a tangible objective product of such a nature that its goodness or quality can be measured. Another reason is that for practical purposes the problem of measuring reading is to discover not what rich and varied meanings the subject draws from the printed page, but rather, how rapidily he can read the material with a sufficient degree of comprehension to get from it the essentials of its meaning.

CHAPTER VII

SCALES FOR DIFFICULTY REACHED

There is another group of school subjects in which the tangible records are fundamentally different from those considered in the last chapter, in that they are not expressed in varying degrees of quality from worst to best. They are school subjects of such a nature that what the pupil does is either right or wrong. Foremost among such subjects are spelling and arithmetic, and in the same class with them, but with less definiteness, are geography, history, and grammar. These are informational subjects and, by the common verdict of society, the information is only valuable if it is accurate and correct. A type of handwriting that is somewhat inferior to another sample may be of almost equal practical value. The same cannot be said of spelling, arithmetic, history, geography, or grammar. Classroom products in these subjects are not judged by better and worse; they are judged by right and wrong; and, because of this fundamental characteristic, they are measured by types of tests and scales different from those that have already been considered.

In spelling, the teacher wishes to know "How hard words can the child spell correctly?"; in arithmetic,

"How hard problems can he do?"; in history, or geography, or grammar, "How hard questions can he answer?" The commonest of all classroom questions is probably that which relates to the difficulty of the work which the child can do correctly. The usual method which has been adopted to answer that question is to prepare a series of tasks carefully graded in difficulty. Those near the beginning of the series are so easy that almost any child in the group can do them successfully; as the series progresses, the tasks become increasingly harder; and near the end of the series they have become so hard that almost no child in the group can do them. In taking the test, the children start at the beginning and work as far as they can. The hardest task successfully completed is taken to measure the degree of accomplishment shown.

Such a test is like a series of hurdles of increasing height, and the object is to discover how high a hurdle the individual being tested can clear. He is given as much time as may be necessary to demonstrate where his limit is. Such tests have the virtues of ease of administration and simplicity of interpretation. Their validity depends upon whether or not the ability being tested really does function with relative independence from the time taken.

The Time Element in Scales For Difficulty

In scales for difficulty, the variable which is measured is the difficulty of the work which the child can do. The difficulty of each successive task is carefully in-

creased and controlled; and the child is allowed to overcome it if he can. The quality of his work must be high enough to be considered as "right." Quality is the passing mark of classroom practice.

Ideally, the scale for difficulty is reserved for the measurement of ability in subjects where the amount of time has no effect upon the score. The independence of time, however, must hold not only for the most difficult problems near the end of the series, but for the easy problems as well; so that on whatever task the child is working, he can answer correctly at once or not at all. Conditions must be such that an extra allowance of time to think about the problem will not help him.

Probably the nearest approach to such a situation among the classroom subjects is that found in the case of spelling. In spelling, the child can write the word correctly, or he cannot. If he does not know what the correct spelling is, there is no way for him to stimulate his memory; and there are few rules to help him. The time element in spelling is easily controlled by dictating the words at regular intervals; but such time control is less essential for spelling scales than for scales in almost any other school subject.

In arithmetic, on the other hand, the time element is especially important. Most children who are old enough to be subjects for testing in the operations of arithmetic are acquainted with its processes, but possess most varying degrees of skill in applying them. Many will fail on a short time arithmetic test,

who, if given unlimited time in which to count, verify, and prove, could ultimately make high scores for their work. In geography, history, and grammar, the same rule holds true. Pupils who are doing poor work in the classroom are frequently able, if given unlimited time in which to verify, review, and attempt to remember, to make high scores in a difficulty test.

The student who devises a scale for difficulty, then, must either present evidence to show that scores for the particular ability he seeks to measure are not affected by differences in time; or he must devise methods by which the amount of time the pupil is allowed to spend on each task within the series may be controlled or recorded.

A common method of handling the time element in tests for difficulty has been to start the children simultaneously, and allow them to work as far through the series as they can, at whatever rates they wish, until, at the end of set period, usually five or ten minutes, they are told to stop. The task which each one has reached and finished correctly when the signal to stop is given is assumed to be the hardest he can do. The assumption is, regardless of the accuracy of his work up to that point, that had he been allowed to continue he would have failed on every task above it. Such an assumption cannot properly be made; for the rates at which children work are not directly proportional to their accuracy. The scores which result from such testing represent genuine difficulty scores for some children, rate scores for

other children, and varying combinations of difficulty, rate, and quality, for still others. They are not comparable, because they do not measure achievements of a single kind. They are conglomerate scores of difficulty, quality, and amount, in unknown and varying proportions. They do not conform to the principle of the single variable.

Difficulty tests have been devised, in which time is controlled or rate recorded for every child at every step. The scores of such testing are comparable, for it is then possible to say, "Of two children who work at the same rate, one is able to do harder work than the other," or "Of two children able to do equally difficult work, one is able to work faster than the other." Such scores may be valid and valuable; but the tests necessary to secure them are so difficult of administration that they are not readily adapted to classroom practice.

Reading Not Measured by Scales for Difficulty

In the studies and experiments made in connection with the present attempts to develop new scales for measuring reading, much labor was devoted to making a reading scale in which the progressive steps consisted of paragraphs of increasingly longer and more unusual words. This was a test of the hurdle type, in which the increasing difficulties in reading consisted of constantly greater difficulties of vocabulary. The difficulties of thought, and the types of response,

required by the instructions contained in the progressive paragraphs, were kept at a constant level.

The attempt to make a practicable working instrument of this scale failed, and had to be abandoned, because it was found that the time consumed, and the difficulty of the task that could be successfully completed, were interdependent variables. If the children were given sufficient time, they were able to follow instructions couched in the most recondite and unusual phraseology. The only way to make the hurdles high enough was to introduce complexities of thought and difficulties of task, which tended to make the test one for qualities other than ordinary reading ability.

Because the test as it stood resulted in nearly undistributed high scores, which could not be taken to represent the varying abilities of the children, experiments were tried in setting a single time limit of five minutes, and seeing which children could reach the highest difficulty in that time. Observation of the children taking the test, and analysis of the resulting scores, showed that the last paragraph read correctly within the time allowed represented for some children the most difficult material they could read, when they had been able to read everything before it correctly. For other children, the latest paragraph read correctly might or might not be taken as the limit of their ability, since they had failed on some of the easier paragraphs below it, and might fail or succeed with apparently equal likelihood on some of the harder paragraphs above. The scores of still other children

were clearly measures not of difficulty but of the rates at which they read, and there was ample evidence that had more time been given, they would have been able to read successfully material much harder than any they were allowed to try.

It was evident that comparable scores in silent reading scales of the difficulty type can only be secured by controlling or recording the time spent by every child on every paragraph. This is readily done where one child is tested at a time; but the administration of such careful timing for large groups of children simultaneously tested presents difficulties which cannot readily be overcome. For these reasons the use of the difficulty scale for the measurement of silent reading was abandoned.

Summary

1. There is a large and important group of school subjects in which quality is not expressed in varying terms from worst to best, but is referred to as either "right" or "wrong." The tradition of society has established for these subjects a passing mark; and products or performances which fail to reach that mark are regarded as of no value. This group includes such subjects as spelling, arithmetic, geography, history, and grammar.

2. The commonest classroom question with regard to these "right and wrong" subjects is "How hard work can the child do correctly?" The usual device adopted in the attempt to answer this question is the test or scale for difficulty reached.

3. Scales for difficulty reached consist of series of tasks carefully graded in difficulty from very easy to very hard. Children start at the beginning and work as far as they can. The hardest task successfully completed is taken to mark the upper limit of the child's ability.

4. In scales for difficulty reached, the quality required is the "right" of classroom tradition. Difficulty is carefully controlled and increased. The variable measured is the child's ability to do work of different degrees of difficulty.

5. Scales for difficulty are best devised for subjects where time is not a controlling factor. Such subjects are rarely found. The independence of achievement scores from time allowed, in any given subject, cannot be assumed; it must be proved; and the burden of proof rests upon the individual who devises the scale.

6. A common method for controlling the time element in scales for difficulty has been to set a definite time allowance for testing and allow children to work as far through the series as they can, at whatever rates they wish, until the signal to stop is given. The task which each one has reached and finished correctly is assumed to be the hardest he can do. This assumption cannot validly be made.

7. Where time is a controlling factor, the scale for difficulty must be so devised that time is controlled or rate recorded for every child at every step. If this is done, and certain other conditions observed, the resulting scores are statistically comparable.

8. Reading is a subject in which the time allowed is of great influence on scores secured through testing. In scales for reading, the time element must therefore be controlled. Since the task of recording rate or controlling time, in a large group of children, for every child, at every step of a difficulty scale, presents nearly prohibitive difficulties of administration, the attempt to make a difficulty scale for reading was abandoned.

CHAPTER VIII

SCALES FOR AMOUNT DONE

In spelling, time is controlled by the rate at which the teacher dictates the words; but in such subjects as geography, history, grammar, and arithmetic, the time factor is of greater importance and is much more difficult to control. Accomplishment in arithmetic depends so directly on the number of minutes allowed that the time element becomes the controlling factor in most scientific tests in that subject. Mr. Stuart A. Courtis of Detroit has devised tests in arithmetic that have been more widely used than any other standard tests in education; and so may be considered as embodying the most generally accepted practice in the measurement of accomplishment in that subject.[1]

The basal Courtis tests consist of series of short problems of a constant level of difficulty, so devised that each one in a given set involves a single fundamental operation in arithmetic. The pupils work during a fixed number of minutes, and the number of examples attempted, and the number of these correctly solved, furnish the measure of the degree of success in the test.

[1] Courtis, S. A. *The Courtis Standard Tests*, Department of Cooperative Research, 82 Eliot Street, Detroit, Michigan.

Tests and scales which follow this general procedure may be classified as measures for amount done. In tests and scales for amount done, the quality requirement is ordinarily the right or wrong passing mark which was discussed in the previous chapter. Difficulty is of a single type, and is maintained at a constant level. A definite time limit is set. The variable which is measured is the amount which the child can do correctly within the time allowed.

It is possible to prepare a series of such scales, which takes the place of a scale for difficulty, and gives the answer to the question, "How hard work can the child do?" In such a series, quality and time are held constant throughout the series; but difficulty, while held constant for each scale, is increased at regular intervals from one scale to the next. The Measuring Scale for Ability in Spelling, by Dr. Leonard P. Ayres, is in reality such a series of scales for amount.[1] All the words given in one test are words of equal difficulty, and the score which the child makes shows the proportion he spelled correctly. He is then given another set of words at the next higher level of difficulty, and so on, progressing from one test to the next until his limit has been reached. Instead of such a series being arranged in ascending steps of a single type of difficulty, it might consist of scales for amount, each for a different type of difficulty. In this case, the series would help answer the question, "Which processes can the child do

[1] Ayres, Leonard P. *Measurement of Ability in Spelling*, Russell Sage Foundation, New York City, 1915.

correctly, and on which does he fail?" The eight Courtis arithmetic tests are a series of this sort.

Scales for amount done, in various forms, furnish the means of discovering which children should be warned to go more slowly, and which encouraged to work faster; who needs to be reproved for careless work, and who should be taught the difficult art of neglecting irrelevant details. Scales for amount, in carefully graded series, can answer such questions as "What does the child know, what can he do, and in what skills or knowledges is he lacking?" In fact, scales for amount, carefully prepared, each for a specific purpose, may be made the instruments for answering most of the questions concerning the abilities and disabilities of children, which the classroom teacher needs to ask. They promise to be, in the future, probably the most useful single instrument for the measurement and diagnosis of ability in classroom subjects.

Reading Measurable by Scales for Amount Done

Reading is a subject which lends itself to measurement by scales for amount done. Scales for amount imply testing material in which the quality of the product, the difficulty of the task, and the time allowed, shall all be amenable to control at constant levels. For practical purposes, the problem of reading involves finding out how rapidly the subject reads the material with a sufficient degree of comprehension to get from it the essentials of its meaning. The

quality demanded by everyday experience is the minimum quality of "good enough to get the essential thought;" and this minimum requirement may therefore be considered as the passing mark, which separates success from failure.

The control of reading difficulty involves, first, selecting a single type or combination of types of difficulty, and securing material which shall conform to that selected type; and second, of preparing that material so that each section of it shall be of the same degree of difficulty as each other section. Equality of difficulty rests upon painstaking analysis of reading ability, recognition of the numerous subsidiary factors which make for success in exercising that ability, and the invention of methods for eliminating such of those subsidiary factors as are undesirable and for controlling the remainder.

In a reading scale for amount, the time element must also be controlled. This may either be done by setting a definite time limit, in which the children do as many tasks as they can; or by asking the children to do a certain number of tasks, and recording the time which each child requires to finish them. For classroom purposes, individual timing of children is difficult to administer, and its results are usually unreliable. The more satisfactory method is to determine the time allowance which will give every child a chance to do something, but prevent any ordinary child from doing everything; a time allowance, that is, which will result in well distributed scores, with neither large numbers of undistributed zero scores—

which would indicate a too meager time allowance—nor large numbers of perfect scores—which would indicate a too generous time allowance. What these proper time limits are for reading can be determined by careful experimentation in the classroom.

In the light of the foregoing considerations, it was decided that the new scale for the measurement of silent reading should be a scale for amount, in which quality, difficulty, and time should be held constant, at carefully determined levels, and the variable to be measured should be the amount of reading the children could successfully do in the time allowed.

Summary

1. Informational subjects in which time is a powerful controlling factor are best measured by tests and scales for amount done. The arithmetic tests devised by Mr. Stuart A. Courtis of Detroit are good examples of tests for amount done. In these tests, quality is held at the "correct" or passing mark of classroom tradition; difficulty is maintained of a single kind and at a constant level for every task within the scale; and a definite time limit is set. The variable measured is the amount the child can do correctly within the time allowed.

2. Tests and scales for amount may be prepared singly or in series. They may be so devised as to furnish information concerning how hard work the child can do, on what process or processes he fails, what special knowledges he lacks, and what sorts of

additional training he should be given. They are well adapted for group testing.

3. Reading is readily measurable by tests and scales for amount done. In such measurement the quality required is reading good enough to get the essential thought. The difficulty of the testing material is confined to a single type, and is maintained at a single level. A time limit is set which shall result in scores fairly distributed between 0 and 100; and the variable measured is the amount of such reading under such conditions that the child can do successfully in the time allowed. The new scale for measuring silent reading, Picture Supplement Scale 1, is a scale for amount done.

CHAPTER IX

SCALES AND TESTS

The scale for handwriting is a measuring instrument used to determine the quality of samples of penmanship. It must be noted, however, that the quality of the writing of an individual depends in part on such matters as the degree of haste or care with which it was produced, and corresponding observations would be equally valid concerning lettering, drawing, or composition. Most people normally have several characteristic handwritings. Their penmanship used in composing a formal note will ordinarily be different from, and better than, their writing in personal memoranda quickly jotted down. In a similar way, the quality of lettering, or a drawing, or a composition, will be in a considerable measure dependent on the conditions under which it was produced, and especially on the amount of time taken to do it.

Moreover, if the results are to be compared in terms of relative merit of achievement, it must be shown that the persons producing them have had equal opportunities to learn the skills or knowledges required. That which is exceptionally good writing or reading or arithmetic work for a third grade pupil may well be poor or mediocre work for an eighth grade

pupil, and that which might receive a grade of 80 or 90 on a scale of 100 in the third grade might deserve rather a grade of 20 or 30 in the eighth. Because of these facts, valid comparisons of varying degrees of achievement can be made only if the individuals producing them are members of a single group or class, and have produced their writings, or other classroom products, under the same carefully controlled conditions.

This is provided for by the formulation of tests to accompany the scales; and while the distinction has not been commonly recognized in educational practice, there are two general rules with respect to scales and tests that may be laid down. The first of these is that no scale can be successfully used for comparing the merit of achievements of the members of a group unless the products measured are the results of standardized tests, uniformly administered to uniform groups. The second rule is the converse of the first. It is that the results produced by standardized tests cannot be validly used for comparing the merit of achievements of the individuals in a group unless the comparisons are made by means of scales. The scale and the test go hand in hand as the inseparable component instruments for the adequate measurement of educational products. The test controls the conditions under which the products are produced; the scale is used to evaluate the results.

It is believed that these two rules are valid, and of the first importance in indicating the steps that must be taken in devising a new method for measure-

ment in education. By derivation from the Latin, the test is the thing with which a sample is taken. It originally meant a little earthenware pot. Later on the term was applied to a little crucible used for testing the fineness of silver, and still later we find it used in almost its original sense as applied to the test tube of the laboratory. The scale, or scala, was a ladder, a long straight article divided into equal steps and readily lending itself to use as a measure.

The scale, as used in the evaluation of educational products, may be thought of as a linear rule extending from the worst to the best, from the product of no merit to that of greatest merit, and having indicated upon it steps or degrees by which intermediate achievements may be gauged. It measures in terms of relativity of merit, showing whether a given product or performance represents an achievement that is halfway along the scale from worst to best, or at a point 90 per cent of the distance from the zero point to the high end, or in some other definite location. The function of the scale is to take the score resulting from the test and interpret it in terms of relative merit.

The result secured from the test itself does not carry with it the interpretation of its value in terms of relative merit because relative merit can be measured only in terms of the whole range for the group in question, subdivided into units of known value, which preferably are equal steps. Since it is the scale and not the test which shows the range of merit of achievement, and marks the steps within that range,

it is by means of the scale that comparisons of the merit of achievement of different individuals must be made.

In brief review the foregoing are the principles underlying the rule that comparisons of merit of achievement in classroom subjects can be made only by using scales to measure products resulting from the use of standardized tests. The corollary of this rule is that every scale must be accompanied by definitely formulated procedures of testing, designed to secure the materials to be measured, and that the results of testing must be measured by scales if they are to be used to compare merit of achievement.

Homogeneous Groups

It will be noted that, in the foregoing discussion, one of the conditions laid down for insuring the validity of the results of measurements has been that the group measured should be homogeneous, in that it should consist of individuals of like degrees of maturity and training. This requirement is so important that it seems to demand special comment. In practice, the term "homogeneous" has to be interpreted somewhat loosely, and has usually been taken to mean that a group tested should consist of children of a given school grade who are approximately of equal maturity and have been given about equal amounts of school training. It is entirely probable that strict adherence to this rule in the construction of scales would increase their value, whether they

were scales for quality of product, for difficulty reached, or for amount done.

Two illustrations may make clearer the need for keeping these distinctions in mind. The first relates to physical measurements, such as those showing the heights of male human beings. Students of anthropometry have long recognized that groups subjected to such measurements must be homogeneous if the results are to be used in careful analyses. For example, one would not get a normal distribution of returns if he measured heights of soldiers in an army, in which some of the divisions consisted of tall Swedish soldiers, while others had been recruited among the shorter peoples of southern Italy. It is still more important that such a group should consist entirely of adult men. If boys and infants were included, the measures might run all the way from about 15 inches to about 75 inches, and their distribution would not at all resemble the familiar bell-shaped outline which will always be found in such measurements for height, weight, or other physical characteristics if the group measured is homogeneous, and which has come to be known as "the normal surface of frequency."

A second illustration may be drawn from spelling, in which words are distributed according to the degree of success that children have in spelling them. Here it is important that the groups tested should be homogeneous with respect to maturity and degree of training, as it was in the case of measurement of height, and for reasons that are fundamentally the same. The range of spelling ability of second grade

pupils is marked at its low end by the ability to spell the simplest words, such as *me* and *do*, and in its higher ranges by ability to spell longer words, like *stop* and *turn*. In the case of eighth grade children, however, these words are not distributed along a base line measuring the spelling ability of the pupils, because they have now become of equal difficulty. The children can spell them all, automatically, and without hesitation. The words *stop* and *turn* are no harder than the words *me* and *do*, and since they are all of no difficulty, they are all of equal difficulty.

In brief outline, these are the reasons why all groups to be measured must be homogeneous with respect to maturity and training, in so far as it is practically possible to obtain such homogeneity. It now seems probable that accepted practice in the development of educational measurements will increasingly recognize the importance of this principle, as it has long been recognized in the older fields of statistical investigation in anthropometry, biometry, and biology.

Tabular Classification of Tests and Scales

In this and the preceding chapters, educational measuring instruments have been classified as tests and scales. Tests have been divided into three groups, according as they are for quality of product, difficulty reached, or amount done. The factors involved in testing have been shown to be three in number, two of which are kept constant, and one of which is allowed to vary, while its variations are noted. The

relationships between these different instruments and factors will be made clearer by the tabular presentation on page 102, which gives the classifications for several of the well-known tests and scales in the fundamental school subjects. In cases where the testing methods have not been fully formulated by the authors, the procedure tabulated is that which has been accepted by common practice.

The tabulation shows that while there are many variations, in general the well known measuring devices conform to the fundamental principles already discussed. Each test makes an attempt to measure a single variable. Each scale is accompanied by a test, which is a set of more or less carefully formulated methods for regulating the conditions under which the product to be measured is produced. Wherever the test is accompanied by a scale, the scale interprets the test data in terms of comparative merit for a particular group.

Author and subject	What test does to				What the scale does	Terms in which final results are given
	Quality of product	Difficulty of task	Amount done	Time consumed		
Thorndike writing	Lets it vary	Holds it constant	Records it	Holds it constant	Arranges samples in defined progressive steps of quality. Assigns credit to each on scale of 4 to 18 for children in general	Quality of product at rate used by pupil
Hillegas composition	Lets it vary	Holds it relatively constant	Records or disregards it	Holds it constant	Arranges samples in defined progressive steps of quality Assigns credit to each, on scale of 0 to 1,000 for children in general	Quality of product regardless of amount
Ayres spelling	Holds it constant	Holds it constant for each test. Increases it by equal steps for consecutive tests	Holds it constant for each step	Holds it constant for each step	Arranges words of equal difficulty in equal progressive steps of difficulty for each grade	Per cent right at each grade difficulty
Courtis arithmetic	Holds it constant	Holds it constant	Lets it vary and records it	Holds it constant		Amount right
Gray oral reading	Lets it vary and records it	Increases it by controlled gradations	Holds it constant for each step	Lets it vary and records it for each step	Arranges passages in defined progressive steps of difficulty Locates scores on scale of credit from 0 to 100 for each grade	Merit of performance at rate used by pupil at each step of difficulty
Thorndike silent reading	Lets it vary and records it	Increases it by controlled gradations	Holds it constant for each step	Lets it vary for each step but does not record it	Arranges passages in defined progressive steps of difficulty Locates scores on scale of credit from 0 to 10 for children in general	Merit of performance regardless of rate

Summary

1. Two rules are formulated. The first is that comparisons of merit of achievement in classroom subjects can be made only by the use of scales to measure products that result from the use of standardized tests.

2. The second rule is that every scale must be accompanied by definitely formulated procedures of testing designed to secure the materials to be measured, and that the results of testing must be measured by scales if they are to be used to compare merit of achievement.

3. The function of the test is to recognize the interdependent factors conditioning the results; to select one as the variable that is to be measured; and contrive, in so far as possible, to keep all the others constant.

4. One of the conditions for securing valid and comparable results in testing is that the groups measured should be homogeneous with respect to the subject under consideration. This principle has long been recognized in the older fields of statistical investigation in anthropometry, biometry, and biology and it seems certain that it will be increasingly accepted as fundamental in the development of measures in education.

CHAPTER X

SCORING THE RETURNS

Through the cooperation of the superintendents, supervisors, and teachers of 21 cities, and two special schools, returns were received giving the scores made by children tested by Picture Supplement Scale 1 in 25 classes in each grade from the third through the eighth. Approximately 30 children were tested in each class. The total number of returns thus secured, excluding such earlier returns as were secured before the experimental material had been revised to nearly its present form, was 4,493. The following list shows the cities in which the school authorities cooperated by carrying on the experimental field work upon which the Picture Supplement Scale 1 is based:

Auburn, New York
Cleveland, Ohio
Columbus, Ohio
Denver, Colorado
Des Moines, Iowa
Detroit, Michigan
East Orange, New Jersey
Grand Rapids, Michigan
Greenwich, Connecticut
Jersey City, New Jersey
Kalamazoo, Michigan
Kansas City, Missouri
Louisville, Kentucky
Lewiston, Maine

Manchester, New Hampshire
Montclair, New Jersey
New York, New York
Newton, Massachusetts
Pittsburgh, Pennsylvania
Springfield, Illinois
Springfield, Massachusetts
Elementary School, University of Chicago
Lincoln School, New York City

The Score Sheet

The returns were made by the teachers on score sheets having the following instructions:

> To the Teacher: Please record in the spaces below the score made by each pupil in the silent reading test. The results are to be used in further modification of the scale, and your cooperation in the work is earnestly asked Write the names of the pupils on the numbered lines at the bottom of this sheet.
>
> In recording scores, please give the number of pictures correctly marked, and tell which ones were wrongly marked and which were skipped. This is a test not of drawing but of reading. Mark as correct any drawing, no matter how crude, which exactly follows instructions. Mark every drawing wrong which does not exactly follow instructions.
>
> Please fill also the blanks at the bottom of the sheet, giving your grade, school, name, etc. Comments on the scales, accounts of special difficulties met by the children, and suggestions for modification will be welcome, and may be written on the back of this sheet.

Below these instructions there were blanks for entering the name of each child, the child's score in paragraphs marked right, the numbers of any paragraphs attempted but incorrectly marked, and lastly, the

numbers of any paragraphs skipped and not attempted.

Equality of Difficulty of Paragraphs

While the scale was being developed every endeavor was made to construct the paragraphs so that they should be of equal difficulty as reading material, of equal difficulty with respect to the instructions they contain, and of substantially equal requirements in the time necessary to read the paragraph and make the mark which supplements the picture accompanying the reading matter. The paragraphs as they appear in the final edition of the scale have been subjected to repeated revisions in order to secure these kinds of equality. In order to test the degree of success resulting from these efforts careful studies have been made of the relation between the number of times that each paragraph has been attempted and the number of times that the instructions have been successfully fulfilled.

In order to increase the significance of this investigation the scale was printed in two different editions. The first of these is the standard edition in which the pictures and paragraphs are presented in the order in which they have been shown in Chapter II. About three-fourths of the children tested used this scale. A second edition was also printed in which the same paragraphs and instructions were employed but in an altered or shifted order with the object of finding out whether the degree of success

among the children was conditioned by the location of the paragraph on the sheet. In both editions the first four paragraphs were left in their original location. In the edition having the shifted order the paragraphs originally numbered from five through 12 changed places with those from 13 through 20, except that paragraph seven of the standard edition became eight in the shifted edition, and the original 16 changed places with 15. The number of returns received from the testing with the paragraphs in standard order was 3,405, while those from the shifted edition numbered 1,088.

After the test was completed, the data from the score sheets showing the numbers of paragraphs wrongly marked and the number of those skipped were aggregated. The results showed for the standard edition and for the shifted edition the entire number of attempts for each paragraph of the test, the number of cases in which the paragraph was correctly marked, and the per cent that those correctly marked were of the number of attempts. Where a paragraph had been skipped, but other paragraphs beyond it had been marked, it was included as a paragraph attempted. The percentages resulting from these comparisons are presented in Table 1.

TABLE 1.—PER CENT THAT CORRECT CASES ARE OF ATTEMPTS FOR EACH PARAGRAPH, WITH THE SCALES PRINTED IN STANDARD ORDER AND WITH THOSE IN SHIFTED ORDER

Paragraph	Per cent correct cases are of attempts	
	Standard order	Shifted order
1	92	88
2	92	88
3	89	90
4	85	88
5	89	93
6	91	86
7	89	92
8	92	93
9	91	89
10	89	95
11	88	89
12	90	88
13	89	91
14	93	91
15	89	93
16	92	89
17	89	90
18	94	87
19	94	96
20	91	95

The percentages presented in Table 1 vary within so moderate a range that they appear to present in both cases evidence that the different paragraphs are of reasonably uniform difficulty for the children. In order to test more fully the relationship between the location of the paragraphs and their difficulty, a computation was made to find the coefficient of correlation between the two sets of data presented in Table 1. Since the coefficient of correlation is a statistical device which measures the degree to which smaller measures in one of two series of paired values tend to be accompanied by smaller measures in the second series, and the degree to which larger mea-

sures in the first tend to be accompanied by larger measures in the second, it follows that if certain of these paragraphs are consistently difficult or easy for the children, these differences should be reflected by the coefficient.

The computation of the coefficient of correlation in this case gave a result of .04 when carried through by the Pearsonian method. This indicates an amount of agreement that is negligible, and a distinctly higher degree of relationship might well have been found from purely chance causes. This evidence, taken in conjunction with the fact that the data of the percentages fall within a distinctly moderate range, indicates that the different paragraphs of the scale are genuinely of substantially equal difficulty for the children.

Tabulation of Data Giving Scores

When the returns were received from the different cities large tabulating sheets were employed ruled with a sufficient number of columns to permit entering the data identifying the records of each city, grade, and individual class, and the number of children having each specified number of paragraphs correctly marked, from nothing to 20 inclusive. The data included returns from 25 classes in each grade from the third through the eighth. The final results are presented in Table 2, which shows the number of pupils marking correctly each specified number of paragraphs in each grade from the third through the eighth.

TABLE 2.—PUPILS IN EACH GRADE MARKING CORRECTLY EACH SPECIFIED NUMBER OF PARAGRAPHS

Paragraphs correctly marked	Grade						Total
	3	4	5	6	7	8	
0	36	7	1	..	.	..	44
1	44	12	7	3	2	1	69
2	88	25	17	11	4	2	147
3	99	36	35	17	9	4	200
4	105	70	45	32	19	9	280
5	77	94	64	34	31	18	318
6	73	102	72	62	51	25	385
7	72	87	92	75	72	48	446
8	37	79	104	84	78	60	442
9	28	59	82	95	82	81	427
10	13	54	73	83	91	79	393
11	13	46	53	72	79	92	355
12	10	31	40	66	70	85	302
13	10	19	30	38	52	63	212
14	1	11	17	36	48	58	171
15	2	8	10	22	29	50	121
16	..	7	7	14	20	28	76
17	..	5	5	5	10	21	46
18	.	3	5	5	5	13	31
19	1	1	2	4	5	4	17
20	..	1	1	5	2	2	11
Total	709	757	762	763	759	743	4,493

Characteristics of Grade Distributions

An examination of the data of Table 2 shows that the distributions of the scores for the six grades are closely similar in character. In each case the returns reveal a wide range of ability on the part of the children. In each case relatively few pupils make very low scores, fairly good scores are far more numerous, and very good ones are relatively rare. The scores show that the test is of sufficient length so that few pupils can correctly complete all 20 paragraphs within the time limit of five minutes.

The progress of the children in ability from grade to grade is reflected by the figures showing the average number of paragraphs correctly marked. These figures are given in the second column of Table 3 below. They show that in the third grade the average performance was one of nearly five paragraphs correctly marked. That of the fourth grade is more than two paragraphs better than this, or just over seven paragraphs. From this point on the average score goes up by approximately one paragraph for each advancing grade. If we omit the fractions from the average scores by grades the data give a performance standard of five paragraphs in the third grade, seven in the fourth, eight in the fifth, nine in the sixth, ten in the seventh, and 11 in the eighth.

TABLE 3.—AVERAGE NUMBER OF PARAGRAPHS CORRECTLY MARKED AND STANDARD DEVIATIONS OF SCORES IN EACH GRADE

Grade	Average number of paragraphs correctly marked	Standard deviations of scores
3	4 90	2 98
4	7.31	3 41
5	8.14	3 37
6	9.33	3.40
7	9 96	3 37
8	11.03	3.24

The data of the third column of Table 3 confirm the conclusions that the six distributions conform in a general way to one single type. These data give the standard deviations of the series for each grade.

They indicate that the amount of spread of the distributions is nearly the same in all the grades.

Summary

1. Data were secured giving the scores made by children tested by Picture Supplement 1 in 25 classes in 23 school systems. The total number of children tested was 4,493. About 750 children were tested in each grade from the third through the eighth.

2. The tests were given from two sets of scales on which the paragraphs were printed in different orders of sequence and a comparison of the two sets of results indicates that the different paragraphs are of nearly equal difficulty.

3. An examination of the scores of the children in the different grades shows that the distributions are similar to each other in character.

4. The average performance of the children in the third grade was approximately five paragraphs correctly marked. That of the fourth grade pupils was about seven paragraphs, and for each higher grade the average performance increased by one paragraph until the average of 11 paragraphs was reached in the eighth grade.

CHAPTER XI

ASSIGNING CREDITS FOR SCORES

The data of Table 2, showing the scores of the pupils in each of the six grades, are of such nature as to suggest that the distributions approximate normal distributions. In each array the cases are most numerous at about the middle of the distribution and taper off from that point in both directions. This tapering off does not exhibit, in the upper grades, any marked skew or tendency to run distinctly farther on one side than on the other. In the data for each grade about as many cases are found below the largest entry in the column as there are above it.

Since the characteristics of these distributions are such as would be found in normal distributions, tests have been made to find how close the approximations really are. Each set of original score data given in Table 2 has been considered as representing a surface plotted along a base line with ordinates representing the number of pupils correctly marking each number of paragraphs. For the purpose of explaining the methods used the data for grade five have been chosen.

Since we know the heights of the ordinates and their locations on the base line, the next problem is

to determine what the ordinates would be if the distribution were a normal one. The first step is to discover what the heights of the maximum ordinate at the average would be for a normal distribution having 762 cases, and a standard deviation of 3.37 as in the case of the fifth grade data.

The computation is worked out by using the theorem that, in a normal distribution, the height of the maximum ordinate is equal to the number of cases divided by the standard deviation, times the square root of 2π. This theorem expresses a relationship between the area of a normal surface of distribution and its maximum ordinate that is not so abstruse as it sounds. The square root of 2π is equal to just over 2.5, so that the theorem really means that the maximum ordinate is equal to the number of cases divided by 2.5 times the standard deviation.

Now in a normal surface of distribution very nearly the entire area is enclosed between the base line, the curve, and ordinates erected at -2.5 and $+2.5$ sigma distance. What the theorem really means then is that in the normal surface the relationships between the base line, the area, and the maximum ordinate are almost the same as they would be in a triangle of the same base and height, and that the altitude is equal to the area divided by one-half of the base. This explanation is, of course, only valid if the base line of the surface be considered as limited to 2.5 sigma distance in each direction, and it disregards the difference between 2.5 as the approximate value of the square root of 2π and the more exact figure of 2.506627.

In the case of the data for the fifth grade the computation shows that if they were normally distributed the maximum ordinate at the average would be approximately 90. From this the heights of the other ordinates may be computed by using a table for the ordinates of the normal curve.

Substituting Normal for Actual Distributions

The next problem is to find out what the scores of each grade would be if they were distributed normally. We know for each grade the whole number of pupils tested, the number marking correctly each specified number of paragraphs, the location of the average performance, and the standard deviation of the whole distribution. If we consider the scores as representing a surface of distribution we can adopt a base line running from the lowest to the highest score and erect ordinates representing the number of pupils correctly marking each number of paragraphs. Since we know the standard deviation of the distribution, we may compute in terms of it the distance of each ordinate from the ordinate which is at the average. Thus, in the case of the fifth grade, the ordinate representing the 104 children who marked eight paragraphs correctly is taken as being at the average. Then, since the standard deviation for the scores of the grade is 3.37, the ordinate representing the 92 pupils who marked seven paragraphs is further to the left than the one at the average and distant from it by 1/3.37 of the standard deviation, or .297 of it. By similar methods the locations of all

the ordinates may be determined in terms of their distances from the average as measured by the standard deviation of the distribution.

Measuring the Amount of Agreement

The methods used and the results found in carrying through these computations are presented in Table 4 and illustrated in Diagram 1. The second column of

TABLE 4.—ORIGINAL DATA FOR GRADE 5 SHOWING PUPILS CORRECTLY MARKING EACH NUMBER OF PARAGRAPHS AND DATA FOR NORMAL DISTRIBUTION HAVING SAME AVERAGE AND STANDARD DEVIATION AS ORIGINAL DATA

Original data		Normal distribution replacing original data			
Paragraph	Pupils	Paragraph from average	Sigma distance	Ordinate	Per cent of cases
0	1	—8	2 38	5	1—
1	7	—7	2 08	10	1
2	17	—6	1 78	19	3
3	35	—5	1.49	30	4
4	45	—4	1 19	44	6
5	64	—3	89	61	8
6	72	—2	.59	76	10
7	92	—1	30	86	11
8	104	0	0	90	12
9	82	1	30	86	11
10	73	2	59	76	10
11	53	3	89	61	8
12	40	4	1.19	44	6
13	30	5	1.49	30	4
14	17	6	1 78	19	3
15	10	7	2 08	10	1
16	7	8	2.38	5	1—
17	5	9	2.67	3	..
18	5	10	2 97	1	..
19	2	11	3 27	..	..
20	1	12	3.56	..	..
Total	762	..	..	756	100
Average	8.1	..	..	8	8
Standard deviation	3.37		..	3 37	3.37

the table shows the number of pupils in the fifth grade who correctly marked each number of paragraphs from none to 20. The next to the last column shows what these numbers would have been if the distribution had been a normal one.

Diagram 1 on page 119 presents the same facts in graphic form. The normal curve is shown by the heavy dotted line, while the actual distribution is represented by the irregular solid line joining the ends of the vertical ordinates. The diagram would represent the facts more accurately if both the normal and the actual distributions were represented by series of upright, contiguous columns, but in that case the two surfaces of distribution would coincide at so many points as to make the diagram far less clear than it is in its present form.

The diagram and the figures of the table show that the agreement between the actual and the normal distribution is close. If both sets of data were represented by diagrams drawn to the same scale and superimposed upon each other, the percentage of the area that would be common to both would be 95 and only five per cent of each would lie outside of this area of agreement. Similar computations have been carried through for all the grades with the results that are presented in Table 5 on page 120.

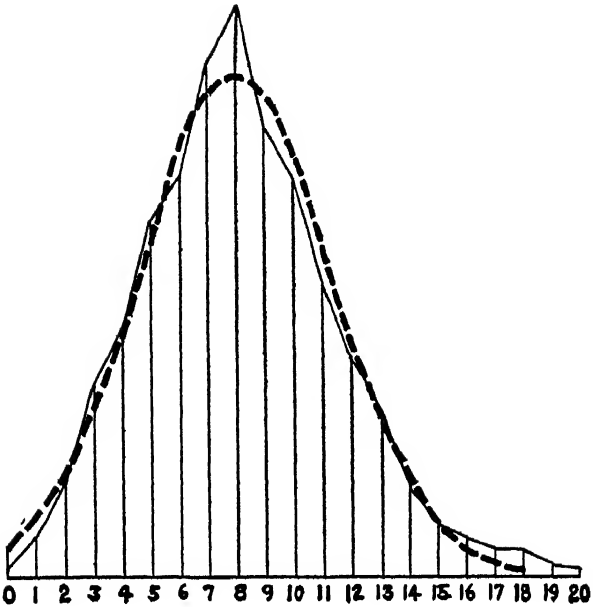


Diagram 1.—The solid line joining the ends of the vertical ordinates represents the actual distribution of 762 pupils in the fifth grade who correctly marked each number of paragraphs from none to 20. The highest ordinate represents the 104 pupils who succeeded in reading eight paragraphs correctly. The dotted line shows the corresponding normal distribution for the same total number of cases and the same standard deviation.

TABLE 5—PERCENTAGES OF COINCIDENCE BETWEEN ACTUAL DISTRIBUTIONS OF SCORES OF PARAGRAPHS CORRECTLY READ BY CHILDREN IN EACH GRADE, AND CORRESPONDING NORMAL DISTRIBUTIONS HAVING THE SAME NUMBER OF CASES AND THE SAME STANDARD DEVIATIONS

Grade	Percentage of Coincidence
3	87
4	93
5	95
6	96
7	96
8	96

In view of the close agreement between the actual and the normal distributions, as set forth by the figures of Table 5, the conclusion has been reached that the normal distributions probably represent the typical results to be expected from using this test more closely than do the actual figures of the original data. In every case the form of distribution is close to the normal, and the departures from the normal appear to be chance variations rather than variations obeying some law or influence of a constant sort.

One Distribution For All Six Grades

The decision to adopt the normal distribution as the most probably valid generalization of the data for the several grades involves the further step of adopting for all six grades a single form of the normal distribution instead of six normal distributions having slightly varying standard deviations. Reference back to the data of Table 2 will show that the standard deviations range in magnitude from 2.98 to 3.47 and it is to be remembered that these measurements are in terms of paragraphs.

Now in a normal surface of distribution so few cases lie further from the center than the ordinates at 2.5 sigma distance in either direction that they constitute less than one per cent of all the cases. This means that the range from a point so low that less than one per cent of the cases is represented, to one so high that again it represents less than one per cent of the cases, is one of five times the sigma distance. Since the sigma distance measured by the standard deviation is in the neighborhood of 3.4 paragraphs, this property of the normal distribution applied in the present case means that there is a range of five times 3.4 paragraphs, or 17 paragraphs from a record so low that less than one per cent of the pupils fail to one so high that less than one per cent of them succeed.

Since the measurement is in terms of paragraphs, it must be made in whole units and not in fractions and for this reason the small variations in the standard deviations may be disregarded and a single standard adopted as approximately representing the typical tendency in all the grades. This standard form of distribution, adopted as typical for all the grades, is the one that has been described. It is a normal distribution covering a span of 17 paragraphs and extending from a point where the failures represent less than one per cent of the pupils, up to one where less than one per cent of them succeed.

It may appear that there should be an exception to this general rule in the case of the third grade, for there the standard deviation is only 2.98 and the

data of Table 2 show that the pupils making the lowest score are more than one per cent of all. The explanation is that the test is somewhat too difficult for the slowest pupils in the third grade and that in this case the distribution is fundamentally similar to those of the other grades, but with its lower end cut off and a resulting concentration of cases at that portion of the array. Nevertheless the evidence indicates that even in this case the normal distribution, adopted as typical for all the grades, is applicable.

The Zero Point and the Upper Limit

The adoption of a single type of distribution for all the grades enables us to translate the scores in terms of values on a scale running from 0 to 100. The zero point on the scale has been taken as being a score so low that it is exceeded by more than 99 per cent of the pupils. The upper limit is a score so high that more than 99 per cent of the pupils fail to reach it. The middle of the scale, at the 50 point, is located in each case at the average performance. There are seventeen paragraphs represented on each grade scale, or sixteen steps of six units each from a point just above zero to one just below 100.

Credits for Scores

The adoption of the criteria that have been described makes possible the construction of Table 6 on page 123 which gives the classroom credits to be awarded for each score in paragraphs correctly marked. The table is in reality a tabular presentation of six scales:

TABLE 6.—CREDIT CORRESPONDING TO EACH NUMBER OF PARAGRAPHS CORRECTLY MARKED IN EACH GRADE

Grade	Number of paragraphs read and marked correctly																				Grade
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
3	0	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	.	..	..	..	..	3
4	0	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	..	..	..	4
5	0	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	.	..	..	5
6	0	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	..	..	6
7	.	0	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	.	7
8	.	..	0	2	8	14	20	26	32	38	44	50	56	62	68	74	80	86	92	98	100	8

one for each of the six grades. By its use all the scores that are ordinarily made by the children in the six upper grades may be translated into classroom credits.

A Simple Statement of Conclusions

It seems worth while to note at this point that the material presented in this chapter is in reality a somewhat elaborate method of reaching conclusions that might be arrived at much more directly. What the present reading scale does is to make available testing material in short, carefully calibrated, units of such a nature that they all test the same kinds of ability, and are all of equal difficulty. The results from using the material indicate that the span of ordinary ability among the children of any one of the upper grades is such that the brightest will correctly mark about 16 paragraphs more than the slowest in the five minutes allowed.

The poorest typical performance is arbitrarily called approximately zero, the best approximately 100, the average is called 50, and the range from poorest to best is divided into 16 equal steps. All of this could be done without using the computations of the normal distribution, but without the added assurance of validity that comes from demonstrating that the results do in fact nearly correspond to the normal distribution.

In general it may be safely noted that school tests could be rendered far better measuring instruments than they have been in the past by developing testing

material made up of short, equal units of material, consistent throughout in nature, and sufficiently numerous so as to provide more work than the brightest pupils could complete within a rigidly controlled time limit. Most useful comparisons could be made from the results of using such tests without any need for introducing elaborate complications in the mathematical treatment of the material.

Distributions From Rate Tests

The fact that the arrays of the scores in the different grades closely conform to the normal distribution is one that calls for additional comment because it represents a result that might not have been expected from *a priori* grounds. The scores result from applying a rate test and it might have been expected that they would produce distributions skewed at the upper end in a manner indicating that equal, progressive increments of ability were reflected by constantly decreasing increments of accomplishment. A result of this sort would naturally be expected in a test largely depending on muscular control and speed of movement, such as writing and copying figures where the fastest performers approach their physiological limit of output. The failure of any such tendency to appear in the present results may perhaps be due to the fact that these reading tests call for manual dexterity in only a minor degree but mainly depend on ability to apprehend correctly the meaning of the printed words.

Summary

1. The distributions of the scores for the six grades are found to conform closely to normal distributions.

2. Since this general conformity to type is found to characterize all six distributions, a single normal distribution has been adopted as representing the best generalized expression of the data.

3. The typical normal distribution substituted for the original data is one running in sixteen steps from a point where less than one per cent of the pupils fail up to a point where less than one per cent of them succeed.

4. Scales are substituted for the original scores so that the results secured by the pupils may be stated in terms of classroom credits running from zero to 100 and giving a credit of 50 for the average performance in each grade.

CHAPTER XII

RELIABILITY OF THE SCALE

The scale for measuring ability in silent reading is an instrument that is designed to secure quickly and definitely certain information concerning school children that would ordinarily come to light slowly and indefinitely, through the regular work of the classroom. The reliability of the scale depends on the degree to which it succeeds in showing, by a sampling process that lasts for five minutes, how much ability any given child has in careful reading. Its practical usefulness will be large if its findings are generally corroborated by the verdict of practical experience; and it will be small if there are many cases in which the record of the pupil is high when tested by the scale, but low in the daily work of the classroom, or vice versa. The scale is in reality a short cut to definite knowledge concerning the ability of children that can ordinarily be secured only through long and varied experience; and because this is its nature, it follows that a true estimate of its reliability can only be made after it has been widely used, and its results have been checked up by the varied records of other classroom work in careful silent reading.

During the experimental work that has developed

the present scale, attempts have been made to test its reliability by using it more than once with the same children, and by comparing its results with those obtained from using other reading tests. While these comparisons do not at all take the place of the results that can only be secured through practical classroom work, they have a certain significance, and are presented in this chapter, as worthy of comment, but not as conclusive.

Five sets of comparisons will be noted. The first of these presents results that were secured in giving the present test twice to the same children. The second set compares the results secured from using the present scale with those obtained from testing the same children with the Kansas Silent Reading Test devised by Dr. F. J. Kelly. The three remaining sets of records present comparisons between results obtained by using the present scale and those from testing the same children with other scales which have been developed in the course of the investigations described in the present volume.

Two Trials of the Same Scale

Through the courtesy of Dr. Harold O. Rugg and the teachers of the Lincoln School, in New York City, tests with Picture Supplement Scale 1 were given to all the children in five grades, on a single morning. The following morning the same test was given to the same children; and the score which each child received on the first test was paired with that which he received on the second. Table 7 shows for

each grade tested the coefficient of correlation between the first and second trials, and its corresponding coefficient of reliability. The probable errors of the correlations are not given because the numbers of cases are in each grade less than 30, and the usual formula for probable error cannot therefore be used.

TABLE 7.—COEFFICIENTS OF CORRELATION AND COEFFICIENTS OF RELIABILITY BETWEEN THE SCORES OF THE SAME PUPILS FOR TWO SUCCESSIVE TRIALS OF PICTURE SUPPLEMENT SCALE 1[1]

Grade	Pupils	Coefficient of correlation	Coefficient of reliability
2	18	.62	.77
3	13	.78	.88
4	19	.62	.77
5	13	.64	.78
6	18	.88	.94
Average	..	.71	.83

[1] Reliability, two trials, $=\frac{2r}{1+r}$ The coefficient of reliability measures the extent to which the amalgamated results of the two trials of a single test would correlate with a similar amalgamated series of results from two other trials with the same test. See, Brown, William, *The Essentials of Mental Measurement*, Cambridge University Press, London, England. 1911, pp 101-2.

It is to be noted in considering the data presented in Table 7 that the coefficients of correlation measure the degree to which the pupils repeated in their second testing the kinds of performances that they made in the first test. The coefficients measure the degree to which children who made good scores in the first test also made good ones in the second test, and conversely, the degree to which those who did poorly the first time also did poorly the second time. When the correlations are fairly high they show that there

was substantial agreement in the results of the two testings, but that this fell short of being complete. These results give us more information with regard to the children than they do with regard to the test. They show us that some children who did well on the first day performed quite differently on the following day; and the same type of statement may be made about those who made poor records on the first trial.

Such results are regularly met with in giving classroom tests, and are familiar to every teacher. They are also found in the more exact measurements of the psychological laboratory. In his book on Mental and Social Measurements, 2d edition, chapter VII, Professor E. L. Thorndike gives, in Table 19, an interesting set of results secured by testing ten individuals 12 successive times with exactly the same test. The results vary so widely from trial to trial that in some instances the subjects made two or three times as high a score in some of their trials as they did in others. Because of these variations, the coefficients of correlation between the scores for the different trials fall far short of being perfect, and vary through a wide range. The coefficient of correlation between the scores of the ten individuals in their first trial and those for their third trial is as low as .36, while that between their first trial and their ninth is as high as .90. The average of the coefficients of correlation between the first trial and the 11 other trials is .65.

The data illustrate in an impressive way the fact that the same individuals exhibit widely varying

capacities and abilities when taking identically the same test on different days. Such variations may constitute in some measure a basis for valid criticism of the test used; but in the main they appear to reflect a real and inevitable variability of human performance. The important fact to remember about such scores is that they may vary from day to day and still be actual true measures of ability on each occasion. Under such conditions the fact that the scores vary from trial to trial does not reflect any inaccuracy or inadequacy of the test or measuring device. The situation is then similar to that encountered when boys are being tested in running to select those who shall represent their school in an athletic meet. The stop-watch is an accurate measuring instrument, but it shows that the same boys run at different rates on repeated trials on different days, and if two sets of the scores be correlated, the resulting coefficients are far from being perfect.

The numbers of cases involved in the data cited by Professor Thorndike are too limited to make the actual coefficients significant; but the figures well reflect the variability that characterizes such measurements. If the data were more numerous, the individual variations would remain, but the range of the correlations would presumably be diminished.

Coefficients of Reliability

The coefficients of reliability given in the last column of Table 7 call for further comment since they are measures that are being used with increasing fre-

quency in educational researches. What Brown's formula really does is to compare the coefficient of correlation between one pair of results from two applications of a test with the coefficient of correlation that would be obtained between one average of scores from two or more testings and another similar average of scores from two or more other testings.

The results furnish a means for determining how many times the test would have to be repeated in order to discover with any desired degree of reliability the relative standings of the different pupils. The method is of limited value because it is impossible to tell whether the correlation between the first two testings is low, average, or high. In the case of the data given by Professor Thorndike, and referred to in the preceding section, the correlations between the various testings of the same individuals with the same test ranged from .36 to .90. If the coefficient of reliability were based on the lowest correlation it would indicate that the results of no fewer than 16 different testings would have to be amalgamated in order to give a reliability coefficient of .90. If it were based on the highest correlation it would indicate that no amalgamation at all would be necessary to produce the same result. It is clear that the method is of limited significance and utility because, when we give a test twice and compute the correlation, we do not know whether the relationship indicated by this single computation is lower or higher than the average of several such correlations would be.

The fundamental assumption behind the method

is that groups repeatedly tested by the same methods will consistently vary to about the same degree, in about the same way, from trial to trial. If this assumption were valid, the correlation between the results of any two trials would be typical of all similar correlations between scores for other pairs of trials. Unfortunately, these assumptions are valid only in moderate degree. Consistency of performance in repeated testings is an essential condition if the coefficient of reliability is to be valid; and, in proportion as such consistency exists, the coefficient becomes unnecessary.

Picture Supplement Scale and Kansas (Kelly) Test

Through the courtesy of Mr. Don C. Bliss, Superintendent of Schools of Montclair, New Jersey, records were secured for three groups of children who were tested by Picture Supplement Scale 1, and had previously been tested by the Kansas Silent Reading Test, devised by Dr. F. J. Kelly. Table 8 shows for these three groups the coefficients of correlation between paired scores in the two tests.

TABLE 8—COEFFICIENTS OF CORRELATION BETWEEN SCORES OF THE SAME PUPILS TESTED BY PICTURE SUPPLEMENT SCALE 1 AND BY THE KANSAS SILENT READING TEST (KELLY). Probable errors of the coefficients of correlation derived by means of the formula $\text{P.E of } r = .67449 \frac{1-r^2}{\sqrt{n}}$

Grade	Pupils	Coefficient of correlation, r	P E. of r	Times r is of its P.E.
3	16	.63	..	..
3	30	.81	.04234	19
6	30	.77	.05060	15
Average	..	.74	..	..

It is noteworthy that the coefficients of correlation between the results from using the present scale and those obtained from the Kelly test are on the average even higher than those obtained from testing the same pupils twice by means of the present scale. Nevertheless, the same comments apply in this case that have been made in the preceding paragraphs which discussed the results presented in Table 7. The Kansas Silent Reading Test has been widely used, but we have as yet few data measuring the degree to which it serves as a trustworthy indicator of the reading ability of the children.

The figures in the fourth column of the table give the probable error of the coefficient of correlation, and indicate that the mathematical chances are even that the true coefficient that would be obtained from repeated trials lies within the range above and below the coefficient actually obtained that is indicated by the amounts stated in the figures of the fourth column. While conservative practice demands that the coefficient of correlation must be several times as large as its probable error to be considered significant, it must be remembered that this is merely a measure of the reliability of the coefficient, and not one of the reliability of the test. In the present instance, the coefficients are many times as large as their probable errors.

Picture Supplement 1 and Continuous Narrative 1

An account has already been given, in Chapter III of this monograph, of five different scales for measur-

ing silent reading which were developed, and the classroom scores of which were secured, in the process of the experimentation which led to the production of Picture Supplement Scale 1. Records were secured for some 1,200 children, most of whom were tested by Picture Supplement Scale 1, and by three of the other five scales developed earlier.

Table 9 shows the coefficients of correlation between scores secured by children on Picture Supplement Scale 1 and scores received by the same children when tested by the Easy Continuous Narrative Scale. This latter scale has already been described at some length in Chapter III. It is a scale which calls for a less careful and more rapid type of reading; the sort of reading, that is, which is ordinarily used when a child reads an easy and interesting story for the pleasure he gets out of it. It is not so much a measure of careful reading as is Picture Supplement Scale 1.

TABLE 9—COEFFICIENTS OF CORRELATION BETWEEN SCORES OF THE SAME PUPILS TESTED BY PICTURE SUPPLEMENT SCALE 1 AND BY EASY CONTINUOUS NARRATIVE SCALE 1

Grade	Pupils	Coefficient of correlation, r	P.E. of r	Times r is of its P.E.
3	174	.61	.0321	19
4	201	.56	.0326	17
5	200	.66	.0269	25
6	176	.47	.0396	12
7	114	.58	.0419	14
8	111	.44	.0516	9
Average	..	.55	..	..

Picture Supplement 1 and Continuous Narrative 2

Table 10 presents coefficients of correlation for the same children as Table 9. In the present case, however, the comparisons are made between scores on Picture Supplement Scale 1 and a longer and considerably more difficult Continuous Narrative Scale. In this instance the coefficients of correlation range from .37 to .74, and in every case the coefficients are many times as large as their probable errors. If the criteria presented by Dr. H. O. Rugg in his book on "Statistical Methods Applied to Education" be adopted, the coefficients of correlation in this and the other tables of this chapter may be regarded as clearly indicating that genuine and controlling relationships are basal to these results. Dr. Rugg concludes (p. 256) that in material of this general sort coefficients of correlation are "marked" when they range from .35 or .40 to .50 or .60, and "high" when they are above .60 or .70.

TABLE 10.—COEFFICIENTS OF CORRELATION BETWEEN SCORES OF THE SAME PUPILS TESTED BY PICTURE SUPPLEMENT SCALE 1 AND BY DIFFICULT CONTINUOUS NARRATIVE SCALE 2

Grade	Pupils	Coefficient of correlation, r	P.E. of r	Times r is of its P.E.
3	174	.74	.0231	32
4	201	.58	.0316	18
5	200	.72	.0230	31
6	176	.74	.0230	32
7	86	.46	.0573	8
8	118	37	.0536	7
Average	..	.60	..	..

Picture Supplement Scale 1 and a Difficult Picture Supplement Scale

The last table in this series, Table 11, presents the correlations for 1,170 children when the scores they made on the new Picture Supplement Scale 1 were compared with their scores when tested by another Picture Supplement Scale, in which the thought, instructions, mechanical presentation, etc., were of the same kind as those in PS-1, but the vocabulary used consisted of much longer words. The two scales measure the same sort of reading ability; but one is at a higher level of difficulty than the other.

TABLE 11.—COEFFICIENTS OF CORRELATION BETWEEN SCORES OF THE SAME PUPILS TESTED BY PICTURE SUPPLEMENT SCALE 1 AND BY A DIFFICULT PICTURE SUPPLEMENT SCALE

Grade	Pupils	Coefficient of correlation, r	P.E. of r	Times r is of its P.E.
3	174	65	.0295	22
4	201	65	.0275	24
5	200	.75	.0208	36
6	206	.71	.0233	30
7	201	.72	.0267	27
8	188	.50	.0369	14
Average	..	.66	..	..

Here, as in the preceding tables, the probable errors are given in the third column and in the fourth there are figures showing that the coefficients of correlation are many times as large as their probable errors. The purpose of comparing the coefficients of correlation with their probable errors is to find out

whether there is good evidence that there is an actual interdependence of the functions being measured. In the present case this means that the object is to find out whether the data indicate that the reading ability of the children is really being measured by the two scales in such a way that poor readers will generally be identified as such by both scales and good readers will be scored as good in each case. The rule is that a coefficient is not considered as good evidence of such an existing correspondence unless it is fairly high and at least three or four times as large as its probable error.

The theory of the use of the probable error is that it indicates the reliability of the data derived from the sample being dealt with. In the present case the statistics which we have relate to a limited group of children, and the question is how far we can trust conclusions based on relatively few cases and from them generalize about the scores to be expected from other children of similar ages and grades.

The rule is that if our cases represent a genuinely unselected random sampling from a much larger group, the reliability of our conclusions can be measured, and will vary as the square root of the number of cases taken. In Table 11 the coefficient of correlation for grade 5 is .75 and the probable error is about .02. This means that if we could test all fifth grade children, instead of merely these 200, and compute the coefficients of correlation for this great number of samples of 200 children each, we should find that one half of our coefficients would be less than

.75+.02, but greater than .75−.02. They would lie between .73 and .77; while the other half would lie outside of these limits. We cannot tell what the true correlation would be, but we can tell something of the limits within which the true correlation would probably vary from the obtained correlation. The statement as to the size of the probable error is a statement of the degree of our ignorance, which grows less as the samples grow larger.

Fortunately, in the present case, the samples are fairly large and taken from a sufficiently varied set of cities so that they may perhaps approximate true random samplings. Since the coefficients of correlation are fairly high, the sizes of the probable errors represent relatively small percentages of probable departure of the true data from the obtained data.

Summary

1. The true test of the reliability of the scale will be found in its degree of utility in the classroom in quickly securing accurate and definite information as to the ability of the children in careful, silent reading.

2. Repeated testings of the same children with the same scale indicate that it operates with a large degree of consistency.

3. There is high correlation between results obtained from testing children with this scale and testing the same children with the Kansas Silent Reading Test.

4. Repeated testings of large numbers of children with this scale, and with other scales developed by the same author, give high coefficients of correlation and low probable errors.

APPENDIX

Chapter III of this monograph contains brief descriptions of the five preliminary scales for measuring silent reading which were developed in the course of the investigations which led to the adoption of Picture Supplement Scale 1. On the pages that follow, a brief statement is given concerning the characteristics of each of these scales. No attempt is made to reproduce them in full, with pictures, method of scoring, etc., but the text of each is included, in the hope that it may be of suggestive value to students of educational measurement.

The five scales were all of the same general plan as Picture Supplement Scale 1; that is, they were all on sheets of paper 11 inches wide and 19 inches long, divided into five columns of four divisions each. All the scales consisted of pictures, and paragraphs of instructions for marking the pictures.

Two Hearing-Reading Scales

The Hearing-Reading scales were printed on both sides. On the front side were 20 pictures. The teacher read aloud instructions for marking each picture with a pencil, and the children, having listened to the instructions, proceeded to follow them. After all the pictures on the front of the sheet had been so marked, in accordance with instructions to which the children had listened, the sheets were turned over. On the back was a similar set of pictures, but in this case the instructions were printed directly beneath each one Instead of listening to the teacher reading aloud, the children read the paragraphs for themselves and, as they finished silently reading each paragraph, they followed the instructions it gave by marking the picture above it with their pencils. Both the

paragraphs read aloud by the teacher and the printed paragraphs that the children read to themselves were graded in ascending difficulty of vocabulary from very simple ones at the beginning to extremely difficult ones towards the end of each scale. Each paragraph on one scale corresponded to another of approximately the same difficulty on its companion scale.

The underlying idea of these hearing-reading scales was to use them first to secure a record of the ability of the child to follow instructions received through hearing them, and then to secure a second record of his ability to follow another set of instructions of equal difficulty, when he had to get their meaning by reading them himself. Both sets of records having been secured, they were to be compared, in order to find out how nearly the child, when reading, could equal the record that he made when he listened and did not have to read.

The material which is reproduced below is the text of the 20 paragraphs printed on the silent reading side of Hearing-Reading Scale 1. It is also the material which was read aloud by the teacher in giving the tests of Hearing-Reading Scale 2. The second series of paragraphs reproduces the material which appeared on the silent reading side of the second scale, and was read aloud by the teacher when the tests with the first scale were given.

Hearing-Reading Scale 1

1. This little girl enjoys a ride on her donkey. She has to hold on when he goes fast but she does not mind that. She would like him to go fast all the time. Look at the stick that is in her hand and with your pencil draw a cord at the end of it to make a whip that she may use if she has need of one.

2. This man and his horse have been having a good time. They have gone a long way very fast, but the strong horse is not tired. He likes to trot and gallop, and now he rears up on his hind legs to show that he wants to jump. Draw a fence in front of him so that he can jump over it and show his master how well trained he is.

cover is on it, for it is easy to hit the man back of you and wound him. Take your pencil and in this picture of the Arab chief draw a round ball on the point of the tall spear so as to keep people who may be struck by it from being badly cut.

4. When people are in a great hurry to write the word NUMBER they sometimes use the short form NO., which means the same thing, to save time and space. The number of this section is 4 and you are now asked to write 4 in the center of the sign placed just over this printed part which tells you what the two letters mean

5. This little boy is wearing his older sister's best hat to make believe he is a girl. He is going along the street to visit his friends and make them laugh. Take your pencil and write the words THIS IS on one side and A BOY on the opposite side of the picture so that people can learn at a glance which he is.

6. This banner tells you that the weather will be fair. If you now look back at the previous picture, you will notice that the child has an umbrella to protect the hat his sister lent him. Put several small crosses upon the handle and body of the umbrella to show that it will not be needed.

7. This tiny chipmunk is poised on a big flat stone watching you with alert eyes and preparing to start away at a moment's notice. With a pencil outline three circles close to the boulder in front of him to represent nuts so that he will know that you intend to be a kind friend and he should not be afraid.

8. Would you prefer a turban instead of a cap? This fellow's is made from several yards of cotton material wound around his head, with the loose ends tucked underneath. Draw a rather short feather standing upright in the folds of the head dress and extending above the head where it can wave in the breeze.

9. These rabbits have been stealing cabbages in the garden. They jumped over the fence, but were startled to find them-

selves in the front yard where everybody could see them. Hastily draw a continuous line around both the naughty little animals to prevent them from escaping until you have a chance to scold them.

10. To make a suitable support for this ornamental and expensive clock take your pencil and with a single deftly directed stroke depict a simple horizontal shelf upon which the clock can remain firmly placed without fear of its being overturned or accidentally injured through a fall.

11. The sleeping children are so comfortable wrapped in their feather coverlets that they will not respond readily to their mother's calling, urging them to hurry dressing. Return therefore to the picture of the clock and write 1,2,3,4,5,6,7, beside it to show that it is striking seven and summoning people to come to breakfast.

12. Here is an ornamental frame decorated with leaf and berry and intended as a plan for embroidery on household linen. You are required at once to print the initial of your last name with your pencil carefully but without hesitation in the vacant square which has been provided for that express purpose.

13. This energetic youth with the long hair is practicing for the approaching big football game. His specialty is long distance kicking, and you are required to hasten to his assistance by drawing a properly inflated pigskin ball soaring up and forward from the blow of his forceful propelling kick.

14. 1920 is leap year, and February has 29 days instead of 28. Upon the accompanying calendar encircle the following dates to indicate that they are all important days: Thursday, February 12, which is Lincoln's birthday, Saturday, the 14th, for Saint Valentine's Day, and Sunday, the 22d, for Washington's birthday.

15. With pencil proceed immediately to blacken the lower section of each bulb in this old-fashioned hourglass as high as the horizontal line, so representing the supply of sand which

originally confined in the upper portion trickles through the intervening aperture and gradually accumulates in the bottom.

16. When this bloodthirsty Zulu throws his murderous spear the unsuspecting foe will be instantly decapitated. With haste turn to the succeeding paragraph and inscribe the admonitory letters DANGER directly before the eyes of the policeman to warn of the impending tragedy and stimulate preventive measures.

17. As you commanded this strenuous officer to extend his jurisdiction to neighboring characters and forestall the attack contemplated by the aboriginal, hastily sketch in his clenched fist a policeman's club, billy, or cudgel with which he may quell the next attack of the unscrupulous potential criminal.

18. Warn the busy public that telephonic communication has been temporarily suspended by inscribing the annoyingly familiar declaration, so often met in similar cases, NOT WORKING to the left of the transmitter and somewhat above the receiver of this practically useless instrument.

19. The pictorial representation of riches frequently assumes the guise here presented. With your pencil contribute a gay touch by portraying in the side of the receptacle an irregular incision through which the laboriously collected dollars may expeditiously escape imprisonment and accrue to the impoverished bystander.

20. The dragon is a mythological beast who symbolizes the extremities of wickedness and is frequently met in the old literature. In an appropriate spot beside this interesting engraving elucidate its meaning to observers who misunderstand its significance by appending the inscription DRAGON.

HEARING-READING SCALE 2

1. This dog sees a cat in the street. He does not like cats and he hates this one. He will watch her and if she comes too near he will bark at her and chase her up a tree. We do not want

him to chase the cat. Take your pencil and draw a rope about his neck so that he can not run after her.

2. This man is what is known as an Eskimo. He lives in the far north where it is cold and there is much snow. When he has to travel he rides on a big sled and is drawn over the ice by a team of dogs. Take your pencil and mark out the whip he carries in his hand so that he will not be able to whip any of his dogs.

3. Someone was reading this big book and left it open on his desk. It should be closed and put back upon the shelves, but the reader does not want to lose his place. With your lead pencil make a light mark on the left hand page so that when he takes the book down again he will know where to begin reading.

4. Look at this poor Indian. His clothes are not thick enough to keep him warm. He is so cold that he has wrapped his blanket close about him. Part of the blanket drags on the ground and will soon become torn and soiled. Put a neat little cross on the end of the blanket which drags on the ground.

5. Have you ever noticed such a strange bird as this? He is not easy to find because he usually is asleep during the daytime and does not leave the dark woods until night begins to fall Take a pencil and make it possible for people to learn what the bird's name is by writing the word OWL beside the picture.

6. This small chap is afraid to start for school without his books. The teacher will scold unless he brings them, but the owl is sitting on them and the little fellow cannot scare him away. Grasp your pencil bravely and cross the owl out of the previous picture with two black lines, so that the child can rescue his belongings.

7. These two flags are used as signals by the men on guard to give warning of probable changes in weather conditions. Write FAIR as a title under the white flag because it indicates pleasant weather and place STORM under the blue one since when it is displayed a storm is coming within twenty-four hours.

8 Glance at these servants who propel between them an unusual covered chair. Have you a notion about the identity of the one riding in the chair? He does not want people to guess his secret In order to prevent curious persons from finding out who is riding inside take your pencil and blacken all the windows.

9. This pictures a Christmas pudding with twigs of holly around the platter and lighted candles at the ends. You can see plump and juicy raisins scattered over the surface. It will be sure to taste delicious. Draw a sharp carving knife thrust in the pudding ready to sever a thick slice for everyone attending the banquet.

10. Hasten to assist this gay and lively young lady, who is having such a delightful afternoon all by herself, by drawing a strong skipping rope in her hands with your ready pencil so that while she is practicing she will enjoy the possession of a real rope instead of an imaginary one.

11. This small maid evidently believes her dress is most correct for outdoor sport. Her neat black slippers and bonnet are especially attractive. With your pencil proceed to blacken the cap of the athletic child in the picture about which you have just finished reading so that she also will be fashionable.

12. The children anticipated Santa Claus' visit and will soon enjoy the attractive presents they find in their stockings. As they will be especially anxious to ascertain the nature of the bundles which are partly concealed picture with your pencil a small hole in one foot through which they can satisfy curiosity.

13. This butterfly has been confined in a warm cocoon, but as summer is approaching he has abandoned his old dwelling and is eagerly exploring the world. Draw a branch on which he can remain comfortably for a while to prevent his newly formed wings from becoming quickly fatigued from unaccustomed exercise.

14. This middle aged gentleman practices diligently in the gymnasium since he is convinced that it is foolhardy to grow

fat. Indicate by sketching a crude circle in front and at a slight elevation the goal toward which his pitching should be unceasingly directed.

15. This dignified old eagle with spreading wings is in a particularly risky position on a smoothly polished sphere. Forestall the likelihood of his being suddenly overturned by drawing with your pencil two stones of moderate size in immediate juxtaposition to the sphere to keep it from rolling.

16. After studying the accompanying banner which indicates approaching cold temperature go back to the seventh paragraph, in which selection signals for FAIR and STORM were reproduced and their significance discussed, and transform FAIR to COLD by inserting a central black square in the white pennant there presented.

17. This cheerfully smiling newspaper worker brings manuscripts to the editorial desk for examination and decision as to final destinations. With pencil picture a narrow tape securely confining this tottering pile, to avoid the inextricable confusion immediately resulting from dropping it.

18. It is certainly interesting to help this noted performer give an exhibition of remarkable dexterity. As he is contemplating the possibility of balancing a stick on his nose increase the difficulty of securing equilibrium by drawing in pencil a moderately small circle representing an orange poised on the extremity of the rod.

19. This curiously made cross is used as a decoration on propagandic literature of the American Tuberculosis Association. As TB is the usual abbreviation of tuberculosis, print the two initials closely contiguous to the celebrated emblem to give suggestions as to the interests of the influential society.

20. Penmanship was once considered a branch of art and incessant application resulted in elaborately undecipherable initials similar to the one here reproduced. With pencil interpret this hieroglyphic by inscribing beside it an unpretentious but neat and legible M.

Two Continuous Narrative Scales

The two continuous narrative scales were on a different plan from those just described. Each scale consisted of a short and interesting story which was divided into 20 sections of equal length, and these sections were then so arranged that the location of each one could only be found by reading the section preceding it. The child was allowed to read for five minutes. Both scales were alike in their general plan; but the second contained one-seventh more material in each paragraph than did the first, and the thought was somewhat harder. The material reproduced below gives the text of both stories, with the paragraphs arranged in the order in which they came if the child found and read each correctly.

It will be noted that in both the Continuous Narrative Scales the paragraphs uniformly carry two instruction thoughts apiece. One of these tells the child where to find the paragraph that follows in the sequence of the story. The other tells him what number he is to write beside the picture of the paragraph when he locates it. Classroom experimentation seems to show that to the child these numbers are a necessary part of the story. In the first scale he writes them down in order to keep track of how many people helped in teaching the little Prince to like books; and in the second he is asked to number the pieces of evidence in the order in which they were presented at the government trial where John testified against the band of spies. The children do not know it, but the fact is that in each scale these numbers run from 1 to 20, and indicate how many paragraphs the child has read up to that point. The result is that the child leaves a record behind him which tells the teacher where he went and what he did. If he has read and marked the story correctly, the highest number written is his score on the test.

Continuous Narrative Scale 1

The Prince's Book

Once upon a time there was a lazy little Prince. He knew how to read, but he did not like to do it. His Father the King was

naturally very much worried, and finally he called Pen and Paper to help him. "They will know how to make my little son like books," he thought, and he sent a messenger to call them. Now Pen and Paper were the first people who tried to help the Prince to like books, so you must find their picture, which is just below this, and write a figure 1 beside it with your pencil. Then go on reading to learn what they said to the King.

"The Prince will like books," said Pen and Paper, "if we write one for him." "Write it, and send it by the Postman," said the King. As the Postman was the second person to help with the Prince's book find his picture in the last column and write 2 beside it. Then go on reading to learn what he did about the book.

"Dear me!" said the Postman, "This is a heavy book. It is full of stories for the little Prince. I must run to the castle and give it to him." If you will look in the third column you will find a picture of the Prince reading his book, and as the book was the third one to help him, write 3 beside it, and read what happened next.

"Will you read me?" asked the Book. "Yes," said the Prince, "I shall begin now and read all night." But his Queen Mother said "NO!" She pointed with her hand. She was the fourth who saw the Prince's book, so hunt for her hand in the fourth column, number it 4, and read what she was pointing at.

"I shall not let you read all night," said the Queen. "When the hourglass says so, you must go to bed," and she pointed to the hourglass. The hourglass was the fifth person to think about the Prince's book, so find its picture at the top of this column and write a 5 beside it. Then read what it said to the Prince.

"If you read fast," said the hourglass, "and if I let my sand run slowly, you will be able to finish one story before bedtime." "Then," said the Prince, "I shall start here at the rabbits' picture." As this picture was the sixth thing that helped the

little Prince, find it in the fourth column, write 6 by it, and read the story.

Two White Rabbits rushed from the woods. Something dreadful had happened. They must tell somebody. Ahead was a big flat rock and on it sat Mr. Chipmunk. As he was the seventh person who helped the Prince, find his picture in the next column and write a 7 beside it. Then read the dreadful news the Rabbits told him.

"Mr. Chipmunk!" cried the Rabbits, "A Dragon is in the Queen's garden. He is knocking down her golden apple trees!" Mr. Chipmunk sat up straight. "Call Bill, the fat boy!" he ordered; and since Bill was the eighth person to help the Prince read, find his picture in the first column, write 8 by it, and read what he did.

"Bill," said Mr. Chipmunk, "Beat your drum! Call all the animals together. We must ask them what to do about the Dragon." Away ran Bill to get his drum. It was the ninth thing which helped the Prince to read, so find it in the second column and write 9 beside it. Then read what happened when the drum sounded.

Bill stood on the rock and beat his drum. "Hurry up-hurry up-up-up-up!" it called, and all the animals for miles around heard it and came running. The first to arrive was Black Horse, who was the tenth one to help the Prince read. Find his picture in the fourth column, mark it 10, and read what they told him

"What's the matter?" neighed Horse. "Dragon's in the Queen's garden!" chattered Chipmunk. "Dear me! What shall we do? Let's ask the Ducks!" Now they were the eleventh set of people trying to help the little prince read, so find them in the second column and write 11 beside them. Then read what the Ducks said.

"What shall we do?" neighed Horse. "Quack!" cried White Duck, and "Quack!" cried Black Duck in a determined way, "Ask Dog. He knows," and they pointed at Dog with their bills. As he was the twelfth person to teach the Prince,

find him in the last column, write 12 beside him, and read what he answered.

Dog hung his head. "I don't know what to do," he said. "Let's ask the Peacocks. They think they're very wise." So all the animals ran to see the Peacocks, who had the thirteenth chance to help the little Prince. Find their picture in column two and number it 13. Then read how they acted when they heard.

"Mr. and Mrs. Peacock, give us advice!" But the Peacocks turned their backs. "Go away!" said Mr. Peacock crossly, "go away!" "Break down their fountain!" neighed Horse, but Donkey, who was the fourteenth person to help the Prince, interfered. Mark his picture in the last column 14 and read why.

"This is no time," brayed Donkey severely, "to talk to silly Peacocks. Dragon is in the garden, and we must get him out. Bees are wise, let's ask them." All the animals rushed to the Beehive; and since it was the fifteenth to help the Prince, find it in the third column, number it 15, and read what happened next.

"Beehive," brayed Donkey, "Where are the Bees?" "Inside," said Beehive, "Listen!" They heard a whisper from inside, "Eat leaves from the magic tree—Buzzz!" Now that tree was the sixteenth thing to help the Prince read, so find it in this column, mark it 16, and read how the animals ate its leaves.

The Rabbits stood on their hind legs and bit off leaves for everyone. As the last piece went down the donkey's throat—FLASH! A Magician with a spear stood before them. As he was the seventeenth person to help the Prince, find him in the second column and write 17 beside him. Then read what he did.

"What do you want?" thundered the Magician. "Please, Sir, Dragon is breaking down the Queen's apple trees!" "He must stop that!" said the Magician, "Slave, appear!" Up from the earth sprang a Savage Warrior whom you must find

in the last column and mark 18. Then read what the Magician made him do.

"Take my spear," ordered the Magician, "and drive the Dragon out of the garden." "Yes, Sir, right away, Sir," and grasping the spear Warrior dashed away to find the Dragon who was the nineteenth person to help the Prince read. Find him in the first column, write 19 beside him, and read what the Warrior did.

Warrior threw his spear straight into Dragon's tail; and off rushed Dragon, out of the garden, through the fields, and over the mountains, with the spear still in his tail. "Never," said he, "will I go back there again!" and it was twenty years before he did. So write 20 under the 19, to show how long it was.

Continuous Narrative Scale 2

How John Saved the Warship

This page is a test to see if you can think as clearly and act as carefully as John Tuxon, who in 1917 discovered a band of spies, saved a United States warship, and appeared as chief witness in a famous trial. Each picture represents one step in the evidence presented at the trial. You must read carefully enough to find each new piece of evidence and number it according to directions given in the paragraphs. This is the story: A circus reached town at midnight and John's parents let him watch it. He walked beside a cowboy riding a broncho. The cowboy's picture appears below. As he was the first of the accused to testify at the trial, make a number 1 beside it and read the paragraph below it.

In the darkness no one paid any attention to the boy. Ahead of the horseman walked a Chinaman. John noticed that he carried a fan behind which he talked in a low tone to the man beside him. The Chinaman is pictured in the third column. As he was later the second on the list of those tried, number him 2. Then go on reading and find out to whom the Chinaman talked next.

The Chinaman fell back beside the cowboy and muttered a few words which John did not catch. Then he strode forward to catch up with a gorgeously attired Mexican who, wearing a large sombrero, strolled along smoking a black cigar. His picture is in the second column and should be marked 3. Then read what John heard the Chinaman saying.

The Chinaman and the Mexican were talking together. John caught the words—"In half an hour, at the fisherman's hut." He was listening for more when suddenly he felt a heavy hand on his shoulder. He looked up into the stern face of a man in the uniform of a French officer. In column 3 is his picture. Number it 4 and read how John felt when the Frenchman caught him listening.

The officer was scowling. "Go home!" he ordered bruskly, and his accent was not that of a real Frenchman. John was frightened. He started back as if to go home, but took a short cut to the circus grounds and hid behind a large Ferris Wheel that was already erected. A picture of his hiding place appears at the foot of this column. Mark it 5 and read what John saw as he watched.

John remembered that the Chinaman had whispered—"In half an hour, at the fisherman's hut." He knew where that hut was and wondered what they could want there. Soon the circus performers reached the grounds and John saw the Chinaman talking to a man in a black coat, white trousers, and tall white cap, whose picture is in the fourth column. Mark it 6 and then read to learn what they did.

Soon the cowboy, Mexican, and French officer strolled away and disappeared in the night. The Chinaman and his companion approached a jolly Irishman with a silk hat and stick who was amusing the crowd by the campfire. A picture of him is in the last column. As the seventh step in the evidence number it 7 and go on reading to find out why the Chinaman was interested in the Irishman.

They whispered to the Irishman and hurriedly followed the others. The Irishman, still laughing, left the crowd,

giving a light tap as he passed on the shoulder of an elaborately robed oriental king whose crown shone in the firelight. His picture will be found in the fourth column. Number it 8. Then read where the conspirators went and what John did when he saw them go.

By this time John was greatly excited. He determined to reach the hut first. Fortunately he knew a good short cut, so that when the first of the conspirators appeared he was safely hidden in the bushes beside the old cottage. A picture of it appears in the first column. This picture was the ninth exhibit in the trial, so number it 9 and read what John discovered in the fisherman's hut.

The men entered the hut and John peeked through the window. Inside was a strange group. Circus life furnished foreign costumes which made good disguises. One fellow wearing turban, jacket, and short loose trousers looked like a harmless beggar boy, but he was not harmless. Find his likeness in the first column and number it 10. Then read the criminal plan he outlined.

He was speaking in low, incisive tones. "The warship sails at four o'clock. It carries large stores of ammunition, a General, his staff, and a battalion of soldiers besides the crew. This is our chance!" A murmur ran through the group. One, costumed as a foreign sailor, laughed. Find him in column 2, number him 11, and read what he said to his fellow conspirators.

"Everything's arranged," said the sailor. "We give the signal. The rest of you go back to the circus and get ready for the performance." He opened the door and, followed by the pretended French officer, slipped into the darkness. John followed them along the shore to the harbor shown in the picture at the top of the last column. Mark it 12 and read what John saw them do.

The men climbed a tall rock near the shore. John saw the gleam of a strong flashlight. It came again. A voice murmured "Submarine!" and then he understood. He remem-

bered a colt in a nearby pasture; and three minutes later John was riding bareback into the night. Silhouette of horse and rider appears in column 3. Mark it 13, and read how John warned the warship of its danger.

It was three miles to the government wharf. As he galloped John wondered if the submarine would enter the harbor or whether it was lying in wait just outside. He heard the sharp challenge of a sentry, and with a gasp of relief knew that he was in time The warship had not yet sailed. Find the picture of the vessel in the fourth column, mark it 14, and read how John saw the captain.

John was led before the captain, who listened to his breathless story with amazement. As soon as John finished the Captain gave a curt order, and five minutes later, with whirring propellers, an airship sailed out over the water, loaded with bombs with which to destroy the submarine. A picture of the airship is in column 2. Number it 15 and read what happened to the spies.

As the airship started the captain turned again to the boy. "The conspirators are hiding, you say, by masquerading as circus performers? Go to the office on the wharf. Tell the sentinel you are acting by my orders. Telephone police headquarters." John obeyed. The instrument he used is in column 4. Write 16 by it and continue reading to learn how John called the police.

At first the police laughed. The boyish voice, strained with excitement, telephoned an incredible message. Finally, convinced that the matter was serious, they roused their men and in swift motor patrol wagons rushed to the circus grounds. The police captain who directed the raid is shown in the last column. Number him 17 and read what happened when they reached the circus.

The police surrounded the circus grounds. At a given signal they broke into the tents and the conspirators were caught red-handed. Still wearing their disguises they were rushed to jail and the entire gang was placed under the guard

of United States soldiers. The sergeant in charge is pictured in the second column. Number him 18 and read what happened after the arrests were made.

Crowds surrounded the jail to see the prisoners. Newspapers printed John's picture and told the story of his adventures. The General gave him a watch, and the officers and sailors whose lives had been saved presented a bag of gold to the young hero. A picture of it appears at the bottom of the last column. Number it 19 and read the letter of presentation.

TO THE BRAVE YOUTH, WHOSE INTELLIGENCE AND COURAGE SAVED A UNITED STATES BATTLESHIP AND THE LIVES OF TWO THOUSAND MEN, THIS GOLD IS GRATEFULLY AWARDED. The bag held twenty hundred dollars in gold. Write 20 by 19 to indicate its contents The exciting trial at which John was chief witness ended with a unanimous verdict of "Guilty."

A DIFFICULT PICTURE SUPPLEMENT SCALE

At the same time that Picture Supplement Scale 1 was being constructed, a companion scale was prepared for testing in the same way at a somewhat higher level of vocabulary difficulty. The thought of each paragraph, and the difficulty of the response required, were maintained at the same constant level as that of the new scale, Picture Supplement Scale 1; but the vocabulary used consisted of longer and more difficult words. The text of the second scale is reproduced below.

DIFFICULT PICTURE SUPPLEMENT SCALE

1. With pencil circumscribe a protective frame about the accompanying silhouette so that it may be utilized, as profiles of Grecian women are so frequently, for a decorative medallion. Care should be taken that the frame is oval in shape and not round, square, or oblong, as might be the case with an ordinary portrait.

2. Draw a cable attached to the basket of this soaring balloon with the free end hanging overboard so that should a tempest arise it would be possible to moor the balloon while

waiting for calmer weather; and before completing the sketch attach a hook to the end of the cable to represent an anchor for use in mooring.

3. The prevalence of influenza and pneumonia leads this woman's physician to order shoes to prevent exposure in cold weather. By blackening the outline drawing represent the addition of foot covering. Do this only to one foot, however, since the sleepy woman has not had opportunity to complete her dressing.

4. As indication that you understand the significance of *No.*, which is a symbol frequently employed in commercial occupations, inscribe 4, which denotes the number of this paragraph, in the space manifestly intended for it above. Make the inscription not in the center but at the right hand extremity of the sign.

5. To shake the poise of this unpleasantly supercilious butler outline a large boulder immediately in his course. This will almost inevitably cause him to stumble and be precipitated headlong; but in order to insure his demoralization draw still another a short distance further in front of him.

6. This fierce dragon moves with incredible rapidity. Below is a ponderous individual who dislikes labor. Draw a rope from the extremity of the dragon's tail through the print to the handlebars of the bicycle in the illustration below so that he will be dragged rapidly along.

7. The dragon will shortly haul this portly gentleman at such excessive speed that a smooth racing track will be necessary beneath his wheels to prevent disaster. Proceed to make one for him but avoid drawing it behind his machine since it is only in the coming ride that he will feel the necessity for one.

8. Santa Claus owns an automobile. Illustrate one possible disadvantage by marking with a cross the point of probable application should any unscrupulous antagonist desire to halt his journey by puncturing the tire of the rear wheel with some sharp instrument, such as a tack or piece of glass.

9. The processes of eating would be facilitated for this uncomfortably veiled Turkish woman if a small incision were made in the veil immediately in front of the organs of mastication; but since this has not been done lead her to discard the veil by making on its surface a large black cross expressive of your disapprobation

10. If a horse were dragging this ponderous log he would be hitched to it by two long leather straps called traces. Rapidly draw two such traces firmly attached to the trunk but not yet harnessed to any willing horse. Draw them fastened to the end of the trunk opposite that on which the men are pulling.

11. This tiny girl is annoyed because her donkey refuses to gallop but instead prefers a leisurely promenade at extremely slow speed. Draw a cord attached to the upper extremity of the stick which she is grasping and extend its length straight over her head horizontally for a short distance to provide an adequate whip.

12 This fellow has a keen desire for tobacco and experiences considerable difficulty in obtaining it. Draw the thick cloud of smoke which, after he has investigated the resources of the neighborhood and procured the desired materials, will shortly be flying upwards over his shoulder and streaming out behind him.

13. The weight this athlete supports over his shoulder is excessive and he will not long be able to maintain it. Therefore extend a horizontal bar immediately underneath the vertical forearm on which the gymnast may rest his elbow. Allow the bar to extend across the entire width of the picture instead of on one side only.

14. Without attempting to produce a drawing of artistic merit hastily sketch a narrow ribbon attached to the upper portion of this beautiful wreath which has been so cleverly constructed of mistletoe and holly; and as you do so take care to provide two short streamers hanging down through the center.

15. This amusingly strenuous urchin demands an additional rod for use in his outdoor sports. Rapidly draw one in his free hand, and as you do so direct its slant so that it reclines over his shoulder and protrudes behind his head instead of paralleling the other rod in front.

16. Carefully encircle the 22d on this calendar to indicate that it is Washington's birthday and a national holiday. The 12th is Lincoln's birthday, and the 14th is the children's festival in memory of St. Valentine. Make a cross on the latter date to show that although a festival day it is not a legal holiday.

17. This energetic youth's specialty is long distance kicking and you are required to depict an incident in a recent hotly contested game by drawing a properly inflated pigskin ball reclining on the turf, instead of soaring lightly upward, to show the lamentable lack of precision which made him miss his kick.

18. Write the numbers in order from 1 through 6 beside this picture of the sleeping children who are so comfortable that they fail to respond to repeated summons. The figures are to indicate that the clock has struck six. Now add the figure 7 to show that it is seven o'clock before they actually rise.

19. This youngster has surreptitiously procured his elder sister's cherished hat, and is parading down the boulevard. As he will not need the umbrella with which he is thoughtfully provided cross it out. To do this make two crosses appropriately located upon the handle and one at the extreme lower tip.

20. Since TB is the usual abbreviation for tuberculosis print the two initials closely contiguous to this celebrated emblem of the American Tuberculosis Association. In doing so be careful to select such locations as will result in the separation of the initials by the intervention of the emblem between them.

INDEX